JMIR Cardio

Electronic, mobile, digital health approaches in cardiology and for cardiovascular health
Volume 10 (2026) ISSN 2561-1011 Editor in Chief: Andrew J Coristine, PhD, Scientific Editor at JMIR Publications, Canada

Contents

Original Papers

Extracting Cardiorespiratory Symptoms From Clinical Notes Using Open-Weight Large Language Models: Method Development and Validation Study (e89480)
Yunbing Bai, Wanting Cui, Joseph Finkelstein. 6

Heart Rate Estimation Using the Galaxy Watch During Maximal Cardiopulmonary Exercise Testing: Cross-Sectional Validation Study (e81917)
Allan Inoue, João Soares, Felipe Antunes-Santos, Alexandre Ferreira, Alberto Gonçalves, João Alcântara, Marcelo dos Santos. 22

Cardiac Telerehabilitation Using a Smartwatch and a Gamified Smartphone App: Single-Arm Pre-Post Feasibility Study (e85808)
Yusaku Arima, Toshiki Kaihara, Megumi Nakamura, Keita Sone, Toshiya Yoshida, Yui Utsugi, Shiori Takizawa, Shunichi Doi, Kei Honda, Yoshikuni Kobayashi, Akira Kasagawa, Yasuhito Kawagoe, Takahiko Kai, Masashi Koga, Takumi Higuma, Yasuhiro Tanabe, Yoshihiro Akashi. 33

Impact of the Cardio-Meds Mobile App on Heart Failure Knowledge and Medication Adherence: Pilot Randomized Controlled Trial (e83022)
Victor Buswell, Emmanuelle Massie, Elena Tessitore, Lisa Simioni, Guillaume Guebey, Hamdi Hagberg, Aurélie Schneider-Paccot, Samaksha Pant, Katherine Blondon, Liliane Gschwind, Frederic Ehrler, Philippe Meyer. 44

Generation of 12-Lead Electrocardiogram Using Smartwatches in Patients With Cardiac Conditions: Clinical Validation Study (e86621)
Mauro Buelga Suárez, Carmen Plaza-Seco, Roberto Matía Francés, Gonzalo Alonso Salinas, Manuel Blanco-Velasco. 62

Mobile Apps for Heart Rate Variability: App Store Search and Content Analysis (e84764)
Eline de Jager, Brian Caulfield, Evgenia Angelidi, Sinead Holden. 77

Smartphone Ownership Among Cardiology Inpatients Pre–COVID-19 and Post–COVID-19 Pandemic: Secondary Analysis of Randomized Controlled Trials (e77141)
Sarah Badger, Praveen Indraratna, Jennifer Yu, James McVeigh, Joseph Magdy, Joan Li, Nigel Lovell, Sze-Yuan Ooi. 91

Development and Internal Validation of the Soluble ST2, Age, and Estimated Glomerular Filtration Rate–Heart Failure Score for Predicting 30-Day Major Adverse Cardiovascular Events After Heart Failure Hospitalization in a Two-Center Vietnamese Cohort: Prospective Cohort Study (e97879)
An Tran, Nguyen La, Van Phan, Bao Nguyen, Hoa Nguyen, Nhung Thai, Tho Pham. 104

XSL•FO
RenderX

Feasibility of Mobile Single-Lead Electrocardiogram Screening for Atrial Fibrillation: Multicenter Cross-Sectional Study (e92799)

Robert Fenske, Denise Rossa, Christian Butter, Felix Muehlensiepen, Jonathan Nübel, Daniel Meretz, Oliver Ritter, Sebastian Spethmann, Rene Mantke, Mario Zerbaum, Nico Draheim, Nicole Lindenberg-Krause, Anja Haase-Fielitz. 118

Involving Families in Cardiac Care Through Remote Patient and Family Management: Focus Group and Journey Mapping Study (e83055)

Julian Houwen, Veronica Janssen, Sara Hondmann, Niels Chavannes, Maaike Kleinsmann, Douwe Atsma, Valeria Pannunzio. 128

Short-Term Arrhythmia Prediction Using AI Based on Daily Data From Implantable Devices: Multicenter Prospective Observational Study (e85841)

Ignacio Fernández Lozano, Joaquín Fernández de la Concha, Javier Ramos Maqueda, Nicasio Pérez Castellano, Rafael Salguero Bodes, F García-Fernández, Juan Benezet Mazuecos, Javier Jiménez Candil, Tomás Datino, Sem Briongos Figuero, Javier Paniagua Olmedillas, Miguel Font de la Fuente, Juan López-Dóriga Costales, Sarai Paz Fernández, Vicente Copoví Lucas. 145

Co-Creation and Usability Testing of a Digital Cardiac Rehabilitation Program (eCardiacRehab): Design and Development Study (e93398)

Oda Nordfonn, Kristin Johnsen Ramstad, Trond Roed Pettersen, Nina Hjertvikrem, Inger-Lise Aamot Aksetøy, Alf Agcaoili Borge, Eva Gerdts, Jan Hesthammer, Torstein Hole, Irene Instenes, Anita Isaksen, Ole Krumsvik, Merethe Landaas, Christian Moltu, Svein Rotevatn, Jan Schjøtt, Cathrine Sommersten, Jan Tryti, Tone Norekvål. 157

Assessment of Wrist-Worn Activity Tracker Accuracy for Measuring Sedentary Behavior in Outpatients With Cardiovascular Disease: Validation Study (e91104)

Koichi Naito, Kazuhiro Izawa, Noriaki Maeda, Yuya Kasai, Haruna Tani, Misaki Kondo, Hajime Iwama. 176

Real-Time Physical Activity Self-Monitoring in Patients Hospitalized for Cardiology Care: Interrupted Time Series Analysis (e92189)

Carlos de Miguel Llorente, Gertjan Sieswerda, Kaoutar Karramass, Charlotte Rommens, Thomas Hoogeboom, Karin Valkenet. 191

Digital Health Use and Interest in Virtual Cardiac Rehabilitation Among Cardiac Inpatients: Cross-Sectional Study (e91767)

Hamza Patel, Jie Ding, Xinyi Zhang, Abel Cherian, Pauline Hyunh, Apurva Sharma, Vennela Avula, Rongzi Shan, Katherine Ornstein, Shireen Khoury, Francoise Marvel, Lena Mathews, Seth Martin, Erin Spaulding. 206

Postangiography Prediction of Renal Replacement Therapy in Acute Myocardial Infarction–Related Cardiogenic Shock: Least Absolute Shrinkage and Selection Operator Nomogram Development and Validation (e79678)

Amin Daoulah, Julia M Ladna, Abdulrahman Arabi, Ahmed Elmahrouk, Shaber Seraj, Hatem M Aloui, Mohammed Alshehri, Wael Almahmeed, Prashanth Panduranga, Nooraldaem Yousif, Amir Lotfi, Mohammed Qutub, Waleed Alharbi, Rajesh Rajan, Mokhtar Abdirahman Kahin, Abdullah Alenezi, Said Al Maashani, Taher Hassan, Jassim Alswuaidi, Fakhreldein Eltaieb, Mubarak Abdulhadi Aldossari, Mohammed Al Jarallah, Abdelmaksoud Elganady, Badr Alzahrani, Abdulrahman Alqahtani, Faisal Omar M Al Nasser, Husam A Noor, Mohamed Alama, Alaa Aldossari, Sultan Al Obaikan, Alsayed Ali Almarghany, Omar Kanbr, Youssef Elmahrouk, Ibrahim A M Abdulhabeeb, Mohammed Balghith, Ahmad S Hersi, Harvey Antony, Adnan Fathey Hussien, Abdulrahman Almoghairi, Mohamed Alama, Mohamed Ajaz Ghani, Hazim Rahbi, Ayman Basardah, Bandar Alshehri, Abdulhadi Alama, Laura AlObaid, Seraj Abualnaja, Tarique Shahzad Chachar, Hassan Khan, Shahrukh Hashmani, Ahmed A Ghonim, Khalid Almerri, Razan Alsofayan, Abdelaziz M Tawengi, Abeer M Shawky, Ali Alshehri, Amr Arafat, Ahmed Jamjoom. 219

Bidirectional Associations Between Autonomic Function, Sleep Quality, and Daily Fatigue and Energy Symptoms in Long COVID: Longitudinal Digital Health Study (e99630)

Nana Aboagye, Mark R Baker, Kenneth Baker, Silvia Del Din. 322

Bidirectional Automated Texting for Cardiovascular Health Among People Living With HIV: Observational Cohort Analysis of a Stepped-Wedge Cluster Randomized Controlled Trial (e69098)

Mechelle Sanders, Donald Harrington, Emma Sass, Marie Thomas, Tameir Holder, Yiqi Tian, Brent Johnson, Andrea Cassells, Jonathan Tobin, Kevin Fiscella. 339

XSL•FO
RenderX

Mindfulness-Based Self-Management Program Using a Mobile App for Patients With Pulmonary Hypertension: Single-Arm Feasibility Study (e79639)
Yuka Takita, Junko Morishita, Sunre Park, Ayumi Goda, Takumi Inami, Hanako Kikuchi, Takashi Kohno, Masaharu Kataoka, Daisuke Fujisawa. 3 5 2

Barriers and Facilitators to Maintaining Physical Activity After an mHealth Intervention for People Poststroke or Transient Ischemic Attack: Reflexive Thematic Analysis (e87493)
Hanna Lagerlund, David Moulaee Conradsson, Charlotte Häger, Lydia Kwak, Lisa Holmlund. 369

Experiences of Patients With Atrial Fibrillation Using Technology to Personalize Self-Care Decision-Making: Interpretive Description Study (e93036)
Kathy Rush, Cherisse Seaton, Alexandra Lukey, Victor Wang, Peter Loewen, Wynne Chiu, Megan Patapoff, Lana Moroz, Jason Andrade. . . 3 9 4

Positive Psychological Well-Being and Determinants of Social Robot Acceptability Among Patients With Heart Failure: Cross-Sectional Questionnaire Study (e83163)
Lisa-Marie Maukel, Karen Bouchard, Jess Fiedorowicz, Kerstin Dautenhahn, Moojan Ghafurian, Thais Coutinho, Caroline McGuinty, Peter Liu, Heather Tulloch. 408

A Massive Open Online Course for Primary Prevention of Stroke: Evaluation of Co-Design Process (e89358)
Aastha Gurung, Seamus Barker, Sabah Rehman, Rachel Climie, Rosanne Freak-Poli, Catherine Burns, Monique Kilkenny, Tara Purvis, Christine Farmer, Brenda Booth, Janet Bray, Dominique Cadilhac, Jan Cameron, Lachlan Dalli, Stephanie Ho, Eleanor Horton, Timothy Kleinig, Lisa Murphy, Mark Nelson, Muideen Olaiya, Amanda Thrift, Seana Gall. 427

Machine Learning Models for Mortality Prediction in Intensive Care Unit Patients With Ischemic Stroke Associated With Intracranial Artery Stenosis: Retrospective Cohort Study (e82042)
Kun Zhang, Ruomeng Chen, Jingyi Yang, Yan Yan, Lijuan Liu, Chaoyue Meng, Peifang Li, Guoying Xing, Xiaoyun Liu. 444

Sleep and Cardiovascular Health Among Women With a History of Hypertensive Disorders of Pregnancy: Pilot Observational Study (e81118)
Anais Hausvater, Elianna Shwayder, Marie-Pierre St-Onge, Linda Kahn, Manuela Plazas-Montana, Lillian Na, Amanda Joa, Leonardo Trasande, Harmony Reynolds. 462

Perceived Potential and Challenges of Supporting Coronary Artery Disease Treatment Decisions With AI: Qualitative Study (e81303)
Khara Sauro, Bishnu Bajgain, Cody van Rassel, Bryan Har, Robert Welsh, Joon Lee. 474

Impact of a Cloud-Based Care Coordination Platform on False Cardiac Catheterization Laboratory Activations and Unnecessary Team Mobilization: Retrospective Cohort Study (e76932)
William Gibson, Dawoud Al Kindi, François Brouillette, Ahmed Aldajani, Omar Chaabo, Yasmine Lachance, Elie Akl, Kshitij Dandona, Giuseppe Martucci, Jean-Philippe Pelletier, Nicolo Piazza, Jeremy Levett, Tomer Moran, Marco Spaziano. 495

Multilingual Video Education for Hospitalized Patients With Myocardial Infarction (EDUCATE-MI): Single-Arm Implementation Study (e82817)
Aileen Zeng, Edel O'Hagan, Sul Kim, Simone Marschner, Mitchell Sarkies, Marina Wassif, Meng Ji, Julie Ayre, Daniel McIntyre, Clara Chow, Aravinda Thiagalingam, Liliana Laranjo. 504

Improving Cardiometabolic Health in Adolescent Girls With Elevated Anxiety: Web-Based Multisite, Pilot, and Feasibility Randomized Controlled Trial (e82781)
Caitlin Edwards, Lisa Ranzenhofer, Jason Lavender, Lauren Gulley, Stephen Aichele, Isabel Thorstad, Natalia Sanchez, Ruby Schrag, Zoe Sinkford, Jill Emerick, Victoria Thomas, Thomas Arnold, Jami Young, Denise Wilfley, Mark Haigney, Lauren Shomaker, Marian Tanofsky-Kraff. 5 1 7

Association Between Type D Personality and Cardiovascular Disease History: Cross-Sectional Study (e79159)
Keren Grinberg, Yael Sela. 540

XSL•FO
RenderX

Group-Based Multitrajectory Modeling Analysis of Joint Trajectories of Blood Pressure and Heart Rate and Their Association With Relative Change in Relative Wall Thickness in Patients With Hypertension: Community-Based Longitudinal Cohort Study (e87091)
Qi Pan, Nanxiang Zhang, Shuo Yang, Chonglong Ding, Jianan Yin, Yi Fei, Xinyan Zou, Yongjun Zheng, Xiang Huang, Jiezhen Feng, Hai Lin, Jinxin Zhang. . . . 546

Health Care Utilization Among Patients With Coronary Artery Disease Using mHealth Wearable Devices: Feasibility Observational Study (e81922)
Indranil Bardhan, Gari Clifford, Mujeeb Basit, DuWayne Willett, Tianjian Guo, Herman Taylor. . . . 566

Effectiveness of a Telehealth Intervention on Functional Status, Anxiety, Depression, and Rehospitalization Among Older Adults Undergoing Coronary Artery Bypass Grafting: Randomized Controlled Trial (e81777)
Jirawan Mala, Usavadee Asdornwised, Kessiri Wongkongkam, Natkamol Chansatitporn, Punnarerk Thongcharoen. . . . 584

Influence of Perioperative Variables on Surgical Site Infection in Patients With Frailty and Prefrailty Undergoing Coronary Artery Bypass Grafting: Retrospective Multicenter Study (e79408)
Mauricio Guerrieri, Omar Mejia, Bianca Meneghini, Natali Rodrigues, William Burato, Maisa Barros, Daniel Wernsbach, Lucas Bartelega, Alexandre Hueb, Fabio Jatene. . . . 600

Digital Health Literacy in Elective Open-Heart Surgery Patients: Cross-Sectional Study (e83454)
Rikke Daugaard, Thomas Maribo, Britt Borregaard, Ditte Skydt, Kristina Bech, Kjersti Skovli, Ivy Modrau. . . . 613

A Novel Convolutional Neural Network–Based Algorithm for Heart Rate Measurement From Ballistocardiography Signals in Diverse Clinical Settings: Observational Study (e71224)
Kumar Chokalingam, Muthukumarasamy Saravanan, Ashish Kaushal, Srishti Rao, Inam Ur Rahman, Ashwathi Nambiar, Mudit Dandwate, Ravi Mahajan, Kunal Sarkar, Gaurav Parchani. . . . 624

Deep Neural Networks for Automatic Atrial Fibrillation Detection Using Long-Term Ambulatory Electrocardiography: Retrospective Diagnostic Accuracy Study (e83714)
Jagdeep Sedha, Jukka Lipponen, Antti Aho, Helena Jäntti, Onni Santala, Tomi Laitinen, Tiina Laitinen, Jari Halonen, Mika Tarvainen, Eemu-Samuli Seljola, Noora Naukkarinen, Olli Rantula, Tuomas Rissanen, Juha Hartikainen, Tero Martikainen. . . . 643

Patient Perceptions of a Digitally Enabled Community Health Worker Intervention: Qualitative Study Among Pilot Trial Participants (e93288)
Jocelyn Carter, Natalia Swack, Narmeen Rehman, Yadira Reyes-Richards, Karen Donelan, Anne Thorndike. . . . 657

The HeartHealth Program: A Mixed Methods Study of a Community-Based Text Messaging Support Program for Patients With Cardiovascular Disease From 2020 to 2024 (e68896)
Brodie Sheahen, Liliana Laranjo, Ritu Trivedi, Tim Shaw, Gopal Sivagangabalan, James Chong, Aravinda Thiagalingam, Sarah Zaman, Pierre Qian, Anupama Indrawansa, Clara Chow. . . . 668

Human vs Artificial Intelligence in Detecting Coronary Occlusions in ST-Elevation Myocardial Infarction: Retrospective Study (e88399)
Aurelia Zimmerli, Marion Dupre, Quentin Liabot, C-Yasmine Sellah, David Meier, Adil Salihu, Victor Weerts, Vladimir Rubimbura, Olivier Roux, Olivier Muller, Olivier Hugli, Pierre-Nicolas Carron, Stephane Fournier. . . . 684

Comparison of Rule-Based Natural Language Processing and Diagnosis Code Approaches to Detecting Bleeding in the Electronic Health Record: Algorithm Validation Within a Retrospective Cohort Study (e87424)
Evan Draper, Jan Anderson, Surbhi Shah, Leilia Brown, Margaret Hanson, Kristin Negley, Jennifer Lutz, Andrea Anderson, John O'Meara, Nicholas Jakowenko, Robert McBane. . . . 692

XSL•FO
RenderX

Viewpoints

Operationalizing AI-Enabled Cardiovascular Biomarkers: A Clinician-Centered Framework for Validation, Governance, and Workflow Integration (e103401)
Dabeluchi Ngwu. 236

Addressing the Challenges in Using Synthetic Data for Health Research: Application to Cardiology (e92930)
Louise Baschet, Sebastien Marque, Stéphane Locret, Jens Jenssen, Vanessa Barbet, Patrick Jourdain. 244

Reviews

Large Language Models in Cardiology: Systematic Review (e76734)
Moran Gendler, Girish N Nadkarni, Karin Sudri, Michal Cohen-Shelly, Benjamin S Glicksberg, Orly Efros, Shelly Soffer, Eyal Klang. 253

Applications of Smart Textiles for Ambulatory Electrocardiogram Monitoring: Scoping Review of the Literature (e74261)
Clarissa Pedrini Schuch, Gabriela Chaves, Bastien Moineau, Sarah Bennett, Meysam Pirbaglou, Edwin Lobo, Milad Alizadeh-Meghrazi. 278

Technologies, Clinical Applications, and Implementation Barriers of Digital Twins in Precision Cardiology: Systematic Review (e78499)
Fatemeh Sarani Rad, Ehsan Bitaraf, Maryam Jafarpour, Juan Li. 299

XSL•FO
RenderX

Extracting Cardiorespiratory Symptoms From Clinical Notes Using Open-Weight Large Language Models: Method Development and Validation Study

Yunbing Bai, MS; Wanting Cui, MA; Joseph Finkelstein, MD, PhD

Arizona Center for Telemedicine and Digital Health, College of Medicine, University of Arizona, 1501 N Campbell Ave AHSL 1156, Tucson, AZ, United States

Corresponding Author:
Yunbing Bai, MS
Arizona Center for Telemedicine and Digital Health, College of Medicine, University of Arizona, 1501 N Campbell Ave AHSL 1156, Tucson, AZ, United States

Abstract

Background: Accurate identification of clinical symptoms and signs (S&S) is essential for the early detection of high-burden cardiorespiratory conditions, including lung cancer, chronic obstructive pulmonary disease, and heart failure. Although symptom data play a central role in diagnostic reasoning and predictive modeling, most S&S information remains embedded in unstructured electronic health record notes, limiting their use in automated phenotyping, surveillance, and clinical decision support. Traditional natural language processing systems struggle with domain variability and contextual nuance in clinical text. Recent advances in large language models (LLMs) offer a promising alternative, yet challenges remain in hallucinations, overinference, and safe deployment. This study evaluated whether locally deployed open-source models could reliably extract cardiorespiratory S&S and map them to *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) codes using optimized prompting strategies.

Objective: This study aims to assess the accuracy of open-source LLMs in extracting explicitly stated cardiorespiratory S&S from clinical notes and mapping them to *ICD-10-CM* codes (R00-R09) and to compare performance across 4 prompt-engineering strategies, including a multimodule LLM framework.

Methods: A total of 593 clinical notes from the MTSamples database were manually reviewed, with 93 notes used for prompt development and comparison using Llama 3.3-70B, and 500 notes used as testing data for the final best prompt setting using both Llama 3.3-70B and gpt-oss-120B. Four prompting conditions were evaluated: (1) instruction-only, (2) *ICD-10-CM* definition–based prompts, (3) assumption-free prompts, and (4) a multimodule LLM framework with postprocessing. Performance was measured using precision, recall, and F_1-score for both S&S extraction and *ICD-10-CM* code generation.

Results: Across all prompt strategies, model performance improved as more structure and constraints were added. Instruction-only prompting demonstrated high recall but poor precision. Incorporating *ICD-10-CM* definitions improved coding accuracy, and assumption-free prompting further balanced precision and recall. The multimodule approach with postprocessing achieved the highest performance during prompt development. On the independent test corpus, entity-level microaveraged evaluation showed that gpt-oss-120B outperformed Llama 3.3-70B in both tasks. For S&S extraction, Llama 3.3-70B achieved a precision of 0.63, a recall of 0.86, and an F_1-score of 0.73, whereas gpt-oss-120B achieved a precision of 0.89, a recall of 0.87, and an F_1-score of 0.88. For *ICD-10-CM* code mapping, Llama 3.3-70B achieved a precision of 0.59, a recall of 0.83, and an F_1-score of 0.69, whereas gpt-oss-120B achieved a precision of 0.90, a recall of 0.84, and an F_1-score of 0.87.

Conclusions: Locally deployed LLMs, when paired with optimized prompting and multimodule orchestration, can accurately extract cardiorespiratory S&S and generate *ICD-10-CM* codes from unstructured clinical notes. This approach increases the level of data safety by enabling on-premises processing without external data transmission and demonstrates strong potential for scalable, domain-adaptive symptom extraction pipelines in biomedical informatics. Future work should expand datasets and evaluate generalizability across clinical domains.

(JMIR Cardio 2026;10:e89480) doi:10.2196/89480

KEYWORDS

natural language processing; NLP; large language models; electronic health records; signs and symptoms; clinical coding; prompt engineering; named entity recognition

XSL•FO
RenderX

Introduction

Understanding the clinical symptoms and signs (S&S) of high-burden cardiorespiratory conditions—such as lung cancer, chronic obstructive pulmonary disease (COPD), and heart failure—is essential for timely diagnosis, risk prediction, and improved patient outcomes. For example, persistent cough, dyspnea, chest discomfort, or hemoptysis can expedite lung cancer detection, allowing treatment at earlier stages when prognosis is more favorable [1-3]. In heart failure, monitoring subtle changes in breathing patterns, exercise tolerance, or peripheral edema supports the early prediction of disease onset and the prevention of acute decompensations [4]. Similarly, the early identification of worsening respiratory symptoms—such as increased breathlessness or sputum production—can help predict and prevent COPD exacerbations, a leading cause of hospitalization and mortality [5].

In clinical practice, symptom data play a pivotal role in both diagnostic decision-making and predictive modeling. Hospital readmission risk models often incorporate structured and unstructured symptom information to identify patients at higher risk for rehospitalization [6]. Predictive algorithms for heart failure onset and COPD exacerbation frequently rely on longitudinal patterns in clinical data—including symptom trajectories—to generate early alerts for clinicians [4,7,8]. Likewise, lung cancer diagnostic pathways integrate symptom reports with imaging and laboratory findings to guide further evaluation [1,3].

Despite their importance, most S&S are documented in unstructured free-text clinical notes rather than structured electronic health record fields. These unstructured data contain nuanced clinical observations valuable for predicting lung cancer, heart failure onset, COPD exacerbation, and hospital readmission risk. However, the complexity, variability, and domain-specific language of clinical text pose significant challenges for traditional natural language processing (NLP) methods [6-8]. Advances in large language models (LLMs) have shown promise in overcoming these challenges, enabling more accurate and timely extraction of clinically relevant information across diverse documentation styles and supporting predictive analytics in high-impact conditions [9,10].

To address these challenges, specialized biomedical NLP models such as BioBERT and BioRAG have been developed to increase accuracy in symptom identification [6-8,11,12]. These models are particularly effective in detecting symptoms associated with chronic conditions and cancers, making them valuable tools for early disease prediction. However, their dependency on task-specific training datasets can restrict adaptability to new medical contexts. Advanced generative models, such as GPT-4o, have shown promise in identifying symptoms across clinical domains and require little pretraining [9,13]. However, most models are for general purposes that are not fully suitable for medical tasks. In addition, while these models are convenient to use, they often run on commercial cloud platforms, which can raise serious privacy concerns when handling sensitive patient data.

In the *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) coding system, there is a dedicated section for S&S, ranging from R00 to R99 (symptoms, signs, and abnormal clinical and laboratory findings). Our previous studies showed that some LLMs, capable of being deployed and operated locally, demonstrated strong performance in extracting S&S and mapping them to *ICD-10-CM* codes within the genitourinary system, specifically in the range of R30 to R39 (symptoms and signs involving the genitourinary system) [14,15]. However, as task complexity increased, model performance became unstable. In this study, we aim to explore the feasibility of using LLMs with different prompt-engineering strategies to accurately extract cardiorespiratory-related S&S from medical text, map the extracted S&S to standardized *ICD-10-CM* codes in the range of R00 to R09 (symptoms and signs involving the circulatory and respiratory systems), and generate structured outputs to support automated evaluation.

Methods

Data Source

This study conducted a comprehensive analysis of clinical notes sourced from the MTSamples database [16], with a specific focus on records related to medical conditions. The initial phase involved identifying notes that contained keywords associated with cardiorespiratory disorders. These notes underwent manual review, during which any entries that did not explicitly describe clinical S&S were excluded. After this process, a final dataset comprising 93 notes was selected. These notes were retained in their original form, without any textual modifications. For evaluation purposes, annotations provided by clinical experts served as the gold standard. We only extracted and annotated S&S with *ICD-10-CM* codes related to the cardiorespiratory system that were able to map to *ICD-10-CM* codes: R00 to R09 [17]. These 93 cardiorespiratory clinical notes served as the training data for prompt development.

An additional 500 expert-labeled clinical notes from the MTSamples database were used as the test dataset. Unlike the training data, the test notes were not limited to the cardiorespiratory domain; instead, they included S&S spanning multiple clinical systems. This design more closely reflects real-world clinical documentation and enables a more realistic evaluation of model performance.

Each clinical note was annotated by 2 individuals including a nurse and a clinical informatician with experience in medical coding. Cohen κ coefficient was used to evaluate the agreement between 2 annotators as previously described. The interrater reliability based on the Cohen κ coefficient was 0.75 (SD 0.06) in this study.

Prompt Engineering and LLM

In this study, we selected 2 open-weight LLMs for evaluation: Llama 3.3-70B and gpt-oss-120B. Prompt development was conducted using Llama 3.3-70B for iterative prompt engineering and method design [18]. We chose Llama 3.3-70B based on our previous studies, in which it demonstrated strong performance in extracting clinical S&S from medical notes [14,15].

XSL•FO
RenderX

After the final prompt framework was established, we evaluated the proposed approach using both Llama 3.3-70B and gpt-oss-120B to assess whether the workflow generalized across different open-weight models [19]. For both models, the temperature was set to 0, and all other hyperparameters were kept at their default settings during evaluation.

We conducted all experiments on a local workstation equipped with a 32-core 2.50 GHz Intel Xeon W7-3565X processor, 256 GB of RAM, and 2 NVIDIA RTX PRO 6000 Blackwell Max-Q graphics cards with a total of 192 GB of graphics processing unit memory. Both models were deployed locally using Ollama, a lightweight and extensible framework for local LLM deployment. Llama 3.3-70B was run using Q4_K_M quantization, whereas gpt-oss-120B was run using MXFP4 quantization [18,19]. No central processing unit offloading was used during inference. The approximate inference speed was 27.88 tokens/s for Llama 3.3-70B and 152.73 tokens/s for gpt-oss-120B. Although gpt-oss-120B has a larger total parameter count than Llama 3.3-70B, its faster throughput on the same hardware is not necessarily unexpected because gpt-oss-120B is a sparse mixture-of-experts model, meaning that only a subset of parameters is active for each token during inference. By contrast, Llama 3.3-70B is a dense model. The difference may also reflect the deployment formats used in this study, as gpt-oss-120B was run in MXFP4 and Llama 3.3-70B in Q4_K_M.

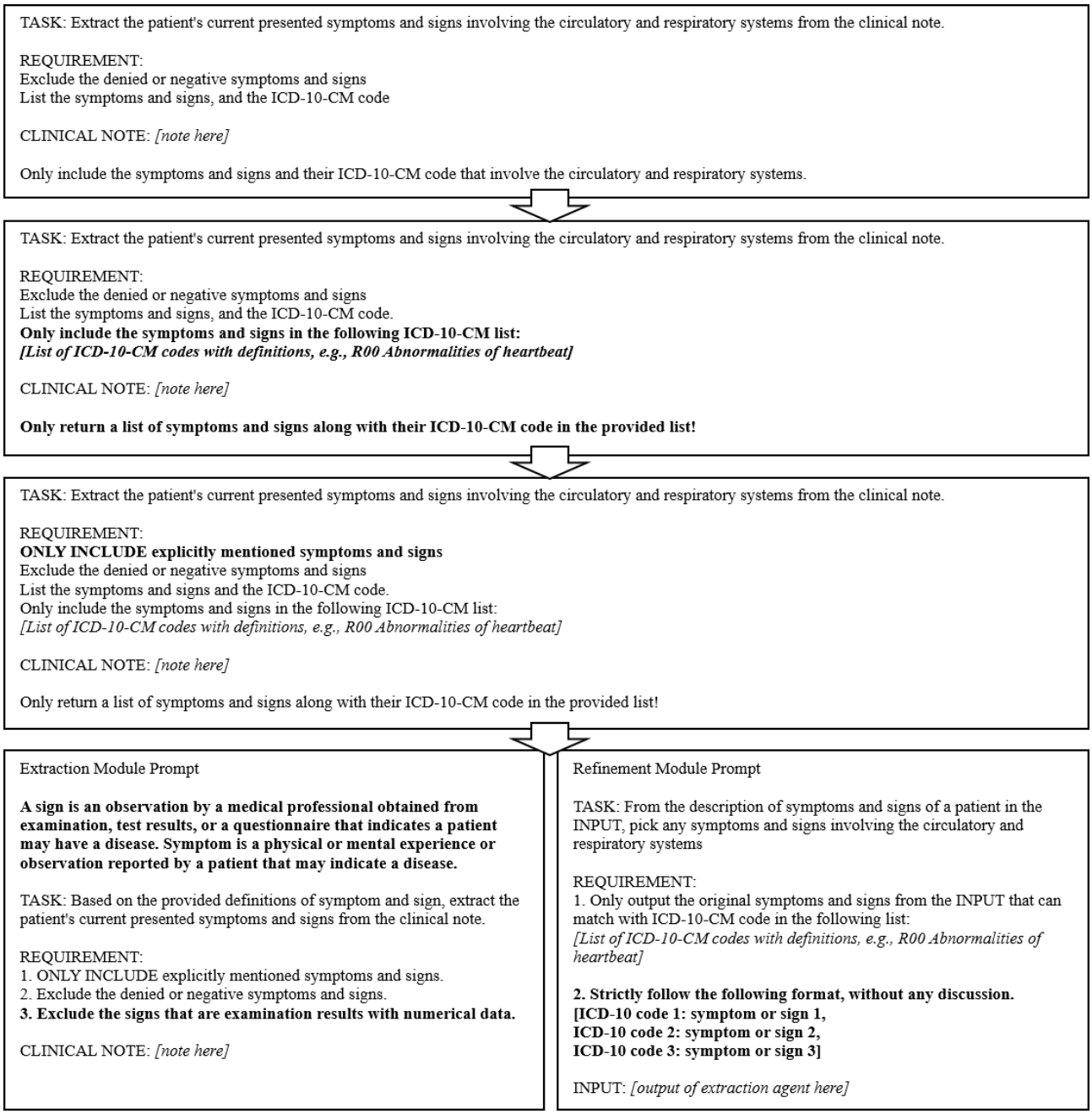
The prompt was designed to accomplish 2 primary tasks: extracting the exact S&S from clinical notes and accurately mapping the extracted S&S to the correct *ICD-10-CM* code groups (R00-R09). To achieve these objectives, multiple prompt structures were tested before selecting an instruction-based prompting approach. The prompt template we applied organized the prompt into 3 distinct sections: "TASK," "REQUIREMENT," and "CLINICAL NOTE." The TASK section clearly defined the objective, instructing the model to extract S&S explicitly mentioned in the clinical notes and associate them with the correct *ICD-10-CM* codes. The REQUIREMENT section provided specific constraints to guide the extraction process. The CLINICAL NOTE section contained the original medical transcription text, from which the model was expected to extract relevant information.

We iterated through 4 generations of prompts, aiming to improve the model's extraction of clinical S&S, as shown in Figure 1. In our initial approach, we provided the task background and goals within the prompt to evaluate the feasibility and general performance of the model without further instructions on extracting S&S or converting them to *ICD-10-CM* codes.

Then, in the second generation, we provided the model with a list of *ICD-10-CM* codes and their definitions within the prompt. To balance the need for completeness while controlling the input length, we limited the *ICD-10-CM* codes to those within the range R00 to R09, specifically including 3-character and 4-character codes (R00-R09 and R00.0-R09.9). These codes corresponded to "symptoms and signs involving the circulatory and respiratory systems" (Table 1). In addition, clinical notes often contained negative S&S, such as "No cough, shortness of breath, fever, or chills." To ensure that only positively reported S&S were extracted, we explicitly instructed the model to exclude any negative S&S from its output. In the third iteration, we aimed to reduce hallucinations by explicitly instructing the model to extract only S&S explicitly mentioned in the clinical notes and to avoid making assumptions about content not explicitly stated.

XSL•FO
RenderX

Figure 1. Overview of the 4 prompt-engineering iterations evaluated in this retrospective methodological study for extracting cardiorespiratory symptoms and signs from unstructured clinical notes. The study used deidentified medical transcription notes from the publicly available MTSamples database. Iterations progressed from instruction-only prompting to *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) definition–based prompts, assumption-free constraints, and finally a multimodule large language model framework with role separation (extraction module and refinement module). Each iteration introduced additional constraints to reduce hallucinations, limit inference, and improve structured *ICD-10-CM* code generation (R00-R09). Bolded elements indicate newly added prompt components at each iteration.



XSL•FO
RenderX

Table . *ICD-10-CM*[a] symptom and sign codes (R00-R09) and corresponding definitions related to the circulatory and respiratory systems that were explicitly provided to the large language models during prompt engineering[b].

ICD-10-CM code	Definition
R00	Abnormalities of heart beat
R00.0	Tachycardia, unspecified
R00.1	Bradycardia, unspecified
R00.2	Palpitations
R00.8	Other abnormalities of heart beat
R00.9	Unspecified abnormalities of heart beat
R01	Cardiac murmurs and other cardiac sounds
R01.0	Benign and innocent cardiac murmurs
R01.1	Cardiac murmur, unspecified
R01.2	Other cardiac sounds
R03	Abnormal blood pressure reading, without diagnosis
R03.0	Elevated blood pressure reading, without diagnosis of hypertension
R03.1	Nonspecific low blood pressure reading
R04	Hemorrhage from respiratory passages
R04.0	Epistaxis
R04.1	Hemorrhage from throat
R04.2	Hemoptysis
R04.8	Hemorrhage from other sites in respiratory passages
R04.9	Hemorrhage from respiratory passages, unspecified
R05	Cough
R05.1	Acute cough
R05.2	Subacute cough
R05.3	Chronic cough
R05.4	Cough syncope
R05.8	Other specified cough
R05.9	Cough, unspecified
R06	Abnormalities of breathing
R06.0	Dyspnea
R06.1	Stridor
R06.2	Wheezing
R06.3	Periodic breathing
R06.4	Hyperventilation
R06.5	Mouth breathing
R06.6	Hiccough
R06.7	Sneezing
R06.8	Other abnormalities of breathing
R06.9	Unspecified abnormalities of breathing
R07	Pain in throat and chest
R07.0	Pain in throat
R07.1	Chest pain on breathing
R07.2	Precordial pain

XSL•FO
RenderX

ICD-10-CM code	Definition
R07.8	Other chest pain
R07.9	Chest pain, unspecified
R09	Other symptoms and signs involving the circulatory and respiratory system
R09.0	Asphyxia and hypoxemia
R09.1	Pleurisy
R09.2	Respiratory arrest
R09.3	Abnormal sputum
R09.8	Other specified symptoms and signs involving the circulatory and respiratory systems
R09.A	Foreign body sensation of the circulatory and respiratory system

[a]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification.*

[b]These codes represent standardized clinical symptom categories used for mapping extracted symptoms and signs in a retrospective analysis of deidentified clinical notes from the MTSamples database. Only 3-character and 4-character *ICD-10-CM* codes within the R00 to R09 range were included to limit the input length of models.

In the fourth iteration, we aimed to prompt the model to produce structured responses that would enable automated evaluation. However, to avoid increasing task complexity and causing unstable outputs, we adopted a multimodule LLM approach in which specialized modules (eg, extraction and refinement) collaborate through role-based interactions to decompose the task and generate standardized outputs—an approach that has been shown to improve reliability and orchestration in complex workflows [20-23].

Specifically, we used 2 modules: an extraction module (EM) and a refinement module (RM). The EM was responsible for identifying all valid S&S explicitly mentioned in the clinical notes. This task consisted of two subtasks: (1) extracting candidate S&S and (2) filtering out false positives. The definitions of S&S from the National Cancer Institute were provided in the prompt, and the model was instructed to include only explicitly stated S&S and to exclude negated or denied findings. During preliminary testing, the model occasionally interpreted examination results—such as blood pressure values—as clinical signs. To mitigate this issue, we further instructed the model to exclude examination findings that contained numerical values.

The RM was responsible for mapping the extracted S&S to *ICD-10-CM* codes in the range R00 to R09. For this module, we provided a predefined list of *ICD-10-CM* codes within this range, along with the required output format. The RM ingested the unstructured plain text output generated by the EM, selected S&S that matched the provided code list, and returned the results in the specified format.

Following model inference, we performed postprocessing and data cleaning on the RM's structured outputs to eliminate assumption-based results. Based on a review of outputs from earlier iterations, we developed a red-flag keyword list (Textbox 1). Any output entries containing terms from this list were flagged as high risk for false positives due to assumptions, inclusion of negated S&S, or hallucinations. All such entries were subsequently removed from the final results.

Textbox 1. Red-flag keyword list used during postprocessing to identify and remove high-risk false-positive outputs generated by the multimodule large language model (LLM) framework. Keywords reflect assumption-based language, negation, inference, or nonexplicit symptom reporting. Any extracted symptom or ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification) code description containing one or more of these terms was excluded from final results.

Red-flag words

- Possible
- No
- Related
- Mentioned
- Not present
- Appears
- Implied
- Normal
- Considered
- Involving
- Explicitly

XSL•FO
RenderX

Evaluation

The model's performance was assessed through a 2-part evaluation for each clinical note. First, the accuracy of extracted S&S was measured to determine how well the model identified only the explicitly stated S&S from the text. Second, the accuracy of generated *ICD-10-CM* codes was evaluated to assess the model's ability to correctly map extracted S&S to the appropriate *ICD-10-CM* categories.

To quantify performance, precision, recall, and F_1-score were used, with human expert annotations serving as the gold standard for comparison. Precision calculated how many predicted S&S were correct, and recall calculated how many ground-truth S&S were recognized. Model outputs were evaluated at the individual clinical note level, and average performance metrics were then calculated across all notes. For *ICD-10-CM* code evaluation, only the first 3 digits of the code were considered to account for variations in granularity. For example, R01, R01.1, and R01.21 were all treated as R01, ensuring consistency in assessment.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Prompt development was conducted using Llama 3.3-70B. Accordingly, performance on the prompt development dataset was compared across prompt iterations to assess how changes in prompt design influenced model outputs. To reduce the risk of overfitting the final framework to either the development data or the model used for prompt engineering, both Llama 3.3-70B and gpt-oss-120B were evaluated on the independent test dataset.

Because the test dataset included notes from multiple clinical systems rather than only cardiorespiratory conditions, many notes contained no annotated cardiorespiratory S&S. In an initial note-level evaluation, notes with both empty ground-truth labels and empty model outputs were treated as correct matches. However, because this macro-averaging approach may inflate overall performance in a sparse dataset, we additionally calculated precision, recall, and F_1-score using standard entity-level microaveraging across the corpus.

For microaveraged evaluation, predicted and reference labels were first reduced to unique sets within each clinical note. True positives were defined as labels present in both the model output and expert annotation for the same note, false positives as labels predicted by the model but absent from the reference set, and false negatives as reference labels missed by the model. These counts were then summed across all notes, and precision, recall, and F_1-score were calculated from the aggregated totals. For *ICD-10-CM* code evaluation, only the first 3 characters of each code were used in this corpus-level analysis. Notes with no gold labels and no predicted labels did not contribute to true-positive counts and were not assigned perfect scores.

In addition to reporting overall performance, we evaluated the model separately for each primary *ICD-10-CM* category within the R00 to R09 range. For this subgroup analysis, both predicted and reference codes were grouped at the 3-character category level (eg, R00, R01, R03, R04, R05, R06, R07, and R09), and precision, recall, and F_1-score were calculated independently for each category. Within each category, true positives represented correctly predicted codes that matched the expert annotation for the same clinical note, false positives represented incorrectly predicted codes assigned to that category, and false negatives represented annotated codes in that category that were not identified by the model. This analysis was included to characterize variation in performance across individual cardiorespiratory *ICD-10-CM* subcategories.

Ethical Considerations

This study did not involve human participants or the collection of identifiable private information. All analyses were conducted using secondary data from the publicly available MTSamples database, which consists of deidentified clinical transcription samples created for educational and research purposes [16]. No patient identifiers were accessed, recorded, or analyzed, and no attempt was made to reidentify individuals. In accordance with journal guidelines, this research does not meet the definition of human subjects research and was therefore exempt from institutional review board review and informed consent requirements. The study was conducted in compliance with applicable ethical standards for research using deidentified secondary data.

Results

Among the 93 human-annotated clinical notes, a total of 168 S&S related to the cardiorespiratory system were identified. On average, each note contained 1.81 (SD 1.09) labeled symptoms. The most frequently observed *ICD-10-CM* category was R06—Abnormalities of breathing, present in 52 notes, followed by R01—Cardiac murmurs and other cardiac sounds (31 notes), R00—Abnormalities of heart beat (26 notes), and R07—Pain in throat and chest (25 notes) [17].

Performance varied across the 4 prompt settings evaluated (Table 2). The instruction-only prompt resulted in high recall but low precision in both tasks, yielding an F_1-score of 0.54 for S&S extraction and 0.41 for *ICD-10-CM* code generation. Introducing *ICD-10-CM* definitions improved both precision and F_1-scores, especially for code generation (F_1-score=0.70), though a slight reduction in recall was observed in extraction. Further improvement was seen when assumption-free constraints were added, increasing F_1-scores to 0.69 for extraction and 0.74 for code generation, with more balanced precision-recall trade-offs.

XSL•FO
RenderX

Table . Performance comparison of 4 prompt-engineering strategies for extracting cardiorespiratory symptoms and signs and generating corresponding *ICD-10-CM*[a] codes (R00-R09).

Prompt setting	Precision	Recall	F_1-score[b]
S&S[c] extraction			
Instruction only	0.38	0.92	0.54
ICD-10-CM code–based symptom extraction	0.51	0.86	0.64
Assumption-free constraints	0.57	0.86	0.69
Multimodule LLMs[d]	0.80	0.94	0.86
Multimodule LLMs with postdata cleaning	0.86	0.94	0.90
ICD-10-CM code generation			
Instruction only	0.29	0.71	0.41
ICD-10-CM code–based symptom extraction	0.58	0.89	0.70
Assumption-free constraints	0.69	0.80	0.74
Multimodule LLMs	0.79	0.96	0.87
Multimodule LLMs with postdata cleaning	0.83	0.95	0.89

[a]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification*.

[b]Precision, recall, and F_1-score were calculated at the clinical note level and averaged across notes. The results are shown separately for S&S extraction and *ICD-10-CM* code generation, demonstrating progressive performance improvements, as additional constraints and a multimodule architecture with postprocessing were introduced.

[c]S&S: symptom and sign.

[d]LLM: large language model.

The highest performance was achieved using a multimodule LLM framework with postdata cleaning. This setting produced an F_1-score of 0.90 for S&S extraction and 0.89 for *ICD-10-CM* code generation, reflecting strong gains in both precision and recall. These results demonstrate the progressive effectiveness of structured prompting and refinement strategies in improving model outputs across clinical NLP tasks.

Overall, the multimodule architecture demonstrated superior performance. Without postprocessing, the model achieved an F_1-score of 0.86 for S&S extraction and 0.87 for *ICD-10-CM* code generation, indicating substantial improvements in both precision and recall. However, the model occasionally produced unstable outputs containing assumption-based errors. Incorporating postprocessing and data-cleaning steps eliminated these assumptions from the multimodule outputs and further improved performance, yielding F_1-scores of 0.90 for S&S extraction and 0.89 for *ICD-10-CM* code generation.

In the test dataset, a total of 1601 S&S were labeled. On average, each clinical note contained 3.20 (SD 2.24) labeled S&S. Among these notes, 69 contained at least one cardiorespiratory symptom or sign, with an average of 1.33 (SD 0.70) cardiorespiratory S&S per note. The multimodule LLM framework with postprocessing was evaluated on this dataset using both Llama 3.3-70B and gpt-oss-120B (Table 3).

XSL•FO
RenderX

Table . Evaluation results of the proposed multimodule large language model framework with postprocessing data cleaning for extracting cardiorespiratory symptoms and signs (S&S) and generating corresponding *ICD-10-CM*[a] codes (R00-R09)[b].

Data and model	Precision	Recall	F_1-score
All testing clinical notes (n=500)			
S&S[c] extraction			
Llama 3.3-70B	0.93	0.97	0.95
gpt-oss-120B	0.98	0.97	0.97
ICD-10-CM code generation			
Llama 3.3-70B	0.94	0.98	0.95
gpt-oss-120B	0.98	0.98	0.98
Testing notes with cardiorespiratory S&S (n=69)			
S&S extraction			
Llama 3.3-70B	0.89	0.83	0.86
gpt-oss-120B	0.97	0.86	0.91
ICD-10-CM code generation			
Llama 3.3-70B	0.90	0.85	0.87
gpt-oss-120B	0.97	0.87	0.92

[a]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification*.

[b]This retrospective evaluation was conducted using an independent test dataset of 500 deidentified clinical notes from the publicly available MTSamples database. The test dataset included notes spanning multiple clinical systems, of which a subset contained cardiorespiratory symptoms. Performance is reported separately for (1) all test notes (n=500) and (2) notes containing at least one cardiorespiratory S&S (n=69). Two locally deployable, open-weight large language models—Llama 3.3-70B and gpt-oss-120B—were evaluated using identical multimodule workflows and postprocessing rules. Precision, recall, and F_1-score were calculated at the clinical note level and macroaveraged across notes. This table demonstrates the robustness and cross-model generalizability of the proposed strategy for cardiorespiratory symptom extraction and *ICD-10-CM* mapping.

[c]S&S: symptoms and signs.

For Llama 3.3-70B, performance on notes containing cardiorespiratory S&S was comparable to that observed in the prompt development dataset. gpt-oss-120B, which is a more recent model with a larger number of parameters than Llama 3.3-70B, achieved overall better performance. This result demonstrates the strong generalizability of the proposed multimodule LLM framework across different model architectures.

Under standard entity-level microaveraging across the full test corpus (Table 4), performance was lower than in the note-level evaluation. gpt-oss-120B outperformed Llama 3.3-70B in both tasks, achieving an F_1-score of 0.88 for S&S extraction and 0.87 for *ICD-10-CM* code generation, compared with 0.73 and 0.69, respectively, for Llama 3.3-70B. These findings indicate that although note-level results were high, corpus-level microaveraged evaluation provided a more conservative estimate of extraction and coding performance.

Table . Standard entity-level microaveraged performance across the independent test corpus (n=500)[a].

Model	Precision	Recall	F_1-score
S&S[b] extraction			
Llama 3.3-70B	0.63	0.86	0.73
gpt-oss-120B	0.89	0.87	0.88
ICD-10-CM[c] code generation			
Llama 3.3-70B	0.59	0.83	0.69
gpt-oss-120B	0.90	0.84	0.87

[a]Precision, recall, and F_1-score were calculated by summing true positives, false positives, and false negatives across all notes after reducing predicted and reference labels to unique sets within each note. The results are shown separately for symptoms and signs extraction and *ICD-10-CM* code generation for Llama 3.3-70B and gpt-oss-120B. *ICD-10-CM* codes were evaluated at the 3-character category level.

[b]S&S: symptoms and signs.

[c]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification*.

XSL•FO
RenderX

Category-specific analysis (Table 5) showed that model performance varied across individual *ICD-10-CM* groups. For gpt-oss-120B, the highest F_1-scores were observed for R05 (cough; F_1-score=1.00), R06 (abnormalities of breathing; F_1-score=0.96), and R04 (hemorrhage from respiratory passages; F_1-score=0.94), whereas lower performance was seen for R09 (other circulatory and respiratory symptoms and signs; F1=0.71) and R03 (abnormal blood-pressure reading; F_1-score=0.73). Llama 3.3-70B showed a similar pattern of stronger performance in more common and clinically explicit categories such as R06 (F_1-score=0.83) and R04 (F_1-score=0.80), but lower performance in R03 (F_1-score=0.33) and R01 (cardiac murmurs and other cardiac sounds; F_1-score=0.40). These results suggest that model accuracy was higher for more explicit and frequently occurring symptom categories, while broader or less frequently represented categories remained more challenging.

For Llama 3.3-70B, 452 (90.4%) of 500 test clinical notes were correctly matched to the expert labels, leaving 48 (9.6%) notes with at least 1 error. Among these 48 error notes, 26 (54.2%) contained hallucinated S&S, 10 (20.8%) included S&S or diagnoses from noncardiorespiratory systems that were incorrectly identified as cardiorespiratory S&S, and 14 (29.2%) missed at least 1 true cardiorespiratory S&S present in the clinical note. These error categories were not mutually exclusive, and some notes contained more than 1 error type. For gpt-oss-120B, 477 (95.4%) of 500 test clinical notes were correctly matched to the expert labels, leaving 23 (4.6%) notes with at least 1 error. Among these 23 error notes, 4 (17.4%) contained hallucinated S&S, 7 (30.4%) included S&S or diagnoses from noncardiorespiratory systems that were incorrectly identified as cardiorespiratory S&S, and 12 (52.2%) missed at least 1 true cardiorespiratory S&S present in the clinical note. As with Llama 3.3-70B, these categories were not mutually exclusive. The most frequent errors reflected assumption-based interpretation, particularly inferring increased or decreased blood pressure from numerical blood pressure readings alone.

Table . Category-specific *ICD-10-CM* code generation performance for each included 3-character cardiorespiratory category in the independent test dataset[a].

ICD-10-CM[b] category	Llama 3.3-70B			gpt-oss-120B		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score
R00	0.42	0.89	0.57	0.75	1.00	0.86
R01	0.33	0.50	0.40	0.67	1.00	0.80
R03	0.40	0.29	0.33	1.00	0.57	0.73
R04	0.73	0.89	0.80	1.00	0.89	0.94
R05	0.75	0.75	0.75	1.00	1.00	1.00
R06	0.70	1.00	0.83	0.96	0.96	0.96
R07	0.53	0.90	0.67	0.77	1.00	0.87
R09	0.47	0.78	0.58	0.85	0.61	0.71

[a]Precision, recall, and F_1-score were calculated separately for R00, R01, R03, R04, R05, R06, R07, and R09 for Llama 3.3-70B and gpt-oss-120B to assess variation in performance across individual cardiorespiratory code groups.

[b]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification*.

Discussion

Principal Findings

We evaluated the ability of Llama 3.3-70B and gpt-oss-120B to extract S&S and map them to corresponding *ICD-10-CM* codes. All *ICD-10-CM* codes were strictly limited to the range R00 to R09 during the manual coding process. For this reason, providing the *ICD-10-CM* codes to the model led to better performance in the *ICD-10-CM* generation task compared to the S&S extraction task. Among the different prompting strategies tested, the multimodule LLMs with postdata cleaning achieved the highest performance across both tasks.

Initially, we only provided basic task instructions to test whether the general-purpose model possessed sufficient medical knowledge from its training data and whether it could correctly interpret medical terminology. The results showed that the model performed well in extracting medical terms. Table 6 presents sample output from 4 different prompt-engineering strategies. While the model was able to identify some S&S terms, it struggled to accurately recognize those related specifically to the cardiorespiratory system. It often failed to distinguish S&S from other medical terms and sometimes misclassified conditions as S&S. For example, the model incorrectly identified “peripheral vascular disease” as a symptom or sign, even though it is a diagnosis. Additionally, it mislabeled terms from other body systems, such as classifying “trace edema at dorsum of feet and ankles” as a cardiorespiratory symptom [14].

XSL•FO
RenderX

Table . Representative outputs generated by the large language model (LLM) under 4 prompt-engineering conditions (instruction-only, *ICD-10-CM* definition-based, assumption-free constraints, and multimodule LLM with postprocessing) for a single deidentified clinical note containing cardiorespiratory symptoms[a].

Prompt setting	Sample output
Instruction only	"Here are the extracted symptoms and signs involving the circulatory and respiratory systems: **Circulatory System:** 1. Ventricular ectopy (I47.1) - PVCs[b] random, run off three beats consecutive 2. History of coronary artery disease (I25.10) 3. Peripheral vascular disease (I73.9) 4. Hypertension (I10) **Respiratory System:** 1. Cough (R05) 2. Shortness of breath (R06.00) 3. Decreased air entry (R09.02) 4. COPD[c] (J44.9) Note that I excluded denied or negative symptoms and signs, as per the requirement. Also, some codes may have additional characters to specify laterality or other details, but I provided the most relevant code based on the information given."
ICD-10-CM[d] code-based symptom extraction	"Here is the list of symptoms and signs involving the circulatory and respiratory systems along with their ICD-10 codes: * Cough - R05 * Shortness of breath - R06.0 (Dyspnea) * PVCs (Ventricular ectopy) - R00.1 (Bradycardia, unspecified is not correct as it's actually ventricular ectopy which falls under other abnormalities of heart beat) - R00.8 * Tachycardia - R00.0 (The patient has a heart rate of 90) * Elevated blood-pressure reading - R03.0 Note: The ICD-10 codes are based on the provided list and may not be comprehensive or entirely accurate without further context or information."
Assumption-free constraints	"Based on the clinical note, the patient's current presented symptoms and signs involving the circulatory and respiratory systems are: 1. Cough (R05) 2. Shortness of breath (R06.0 Dyspnea) The following ICD-10 codes can be assigned: • R05 (Cough) • R06.0 (Dyspnea) Note: PVCs (Premature Ventricular Contractions) are mentioned in the note, but they do not have a direct ICD-10 code match from the provided list. However, it is related to the cardiovascular system and can be associated with abnormalities of heart beat, but since there's no explicit mention of symptoms like palpitations or tachycardia that directly correlate with the PVCs as experienced by the patient, we cannot include R00 codes based on the given instructions."
Multiodule LLMs with post data cleaning	

XSL•FO
RenderX

Prompt setting	Sample output
Extraction module	"Based on the provided definitions of symptom and sign, here are the patient's current presented symptoms and signs extracted from the clinical note: **Symptoms:** 1. Coughing episodes for about two weeks on and off 2. Shortness of breath 3. Neck pain 4. Arthritis (patient is positive for arthritis) **Signs:** 1. PVCs (premature ventricular contractions) random on monitor 2. Decreased air entry in lungs 3. S4 at the apex (heart sound) 4. Cough (mentioned under REVIEW OF SYSTEMS: RESPIRATORY) Note that I have excluded: * Denied or negative symptoms and signs (e.g., no chest pain, no palpitations, etc.) * Signs with numerical data (e.g., heart rate of 90, blood pressure of 187/105, etc.) as per the requirement."
Refinement module	R05: Coughing episodes for about two weeks on and off, R06.0: Shortness of breath, R00.1: PVCs (premature ventricular contractions), R01.1: S4 at the apex (heart sound), R06.8: Decreased air entry in lungs

[a]The examples illustrate differences in hallucination, inference, symptom specificity, and *ICD-10-CM* mapping accuracy. *ICD-10-CM* code evaluation in this study was performed at the 3-character category level; therefore, subcode-level differences were not counted as errors.

[b]PVC: premature ventricular contraction.

[c]COPD: chronic obstructive pulmonary disease.

[d]*ICD-10-CM*: *International Classification of Diseases, Tenth Revision, Clinical Modification.*

The Llama model could generate *ICD-10-CM* codes, suggesting that the coding system was included in its training data by Meta. However, hallucination remained a concern, as the model occasionally produced nonexistent codes. Moreover, some medical terms lacked clear boundaries between symptoms, signs, and diagnoses. In the *ICD-10-CM* system, the R00-R99 range is designated for S&S, but certain valid S&S still had to be mapped to codes outside this range. Due to our labeling strategy, which only included S&S mapped to R00-R09, the model's performance appeared limited. Nonetheless, the model demonstrated its ability to extract medical concepts and associate them with *ICD-10-CM* codes, warranting further investigation.

We then provided a list of *ICD-10-CM* codes and their definitions related to the cardiorespiratory system with the prompt designed to minimize hallucinations and improve the model's ability to identify relevant S&S. Model performance improved significantly and was able to extract most cardiorespiratory-related S&S accurately. Furthermore, the model exhibited the ability to draw inferences based on clinical context, which contributed to high recall. However, this introduced new challenges, including overinterpretation and assumption-based errors, resulting in lower precision. For instance, in one note that mentioned "congested respirations," the model inferred and generated "cough" as a symptom, justifying it with the interpretation: "congested respirations and mild crackles are present, which can be related to cough (R05)." This example illustrates the model's tendency to infer S&S based on contextual associations rather than relying solely on explicitly stated information, resulting in false positives.

To reduce overinterpretation, we refined our third prompt to instruct the model to extract only explicitly stated S&S. This adjustment improved precision, as a higher proportion of the identified terms was correct, while recall declined only slightly.

In the fourth prompt, we enforced a standardized output format that included only the extracted S&S along with their corresponding *ICD-10-CM* codes. This structured format was designed to support an automated evaluation pipeline. However, imposing all constraints within a single model introduced complexity. The model was expected to complete 6 distinct tasks simultaneously: extract S&S, restrict to the cardiorespiratory system, map to *ICD-10-CM* codes, exclude negated S&S, include only explicitly mentioned ones, and format the output accordingly. Relying on a single model led to occasional errors, such as missing explicitly mentioned S&S or including inferred ones, thereby reducing overall accuracy.

The multimodule LLM approach addressed this issue by distributing subtasks across specialized models, leading to more stable and accurate outputs. Additionally, by including definitions of S&S within the prompt, the model gained a clearer understanding of the extraction task. However, through further testing, we observed that the model often misclassified numerical examination results, such as blood pressure readings, as S&S. To address this, we explicitly instructed the model to

XSL•FO
RenderX

exclude examination findings with numerical data. This comprehensive strategy resulted in the highest performance.

Comparison With Prior Work

Prior work on clinical text analysis has largely focused on downstream prediction rather than explicit symptom extraction. Transformer-based models such as ClinicalBERT and Med-BERT have demonstrated strong performance in modeling clinical notes for outcomes such as hospital readmission and disease risk prediction; however, they do not explicitly extract S&S or map them to standardized coding systems such as *ICD-10-CM* [6,8,11,12]. Applying traditional machine learning models to our task would require large amounts of labeled training data for both named entity recognition and entity linking, substantially increasing data requirements, time investment, and labor burden.

More recent studies have applied LLMs to structured information extraction using prompt engineering. For example, prompt-based GPT approaches have been used to extract clinical factors from pathology or radiology reports with high accuracy and improved efficiency compared with manual abstraction [9]. However, these studies typically relied on proprietary, cloud-based models and did not systematically address hallucination or over-inference when processing narrative clinical notes. Hybrid retrieval-augmented generation pipelines have also been proposed to improve precision in extracting specific clinical attributes, such as substance use, by constraining model input to relevant text segments [13]. While effective, these approaches introduce additional system complexity and external dependencies.

In contrast, our study demonstrates that a locally deployed, open-source LLM, combined with assumption-free prompting and a multimodule framework, can accurately extract explicitly stated cardiorespiratory S&S and map them to *ICD-10-CM* codes. This work extends prior research by emphasizing hallucination control, standardized coding, and privacy-preserving deployment, highlighting the feasibility of open-source LLMs for reliable clinical symptom extraction. Future model development should adhere to principles of responsible artificial intelligence apps in medicine [24]. Accurate extraction of symptoms from clinical notes can facilitate artificial intelligence–assisted clinical decision support embedded in electronic health records [25,26].

Limitations and Future Directions

Although the multimodule framework with postprocessing achieved the best overall performance, it did not fully eliminate subcode-level imprecision in *ICD-10-CM* mapping. For example, in Table 6, the refinement module mapped premature ventricular contractions to R00.1, which denotes unspecified bradycardia and is clinically discordant with premature ventricular contractions. This, therefore, represents a clinically incorrect mapping error. However, because our evaluation considered only the first 3 digits of the *ICD-10-CM* code, this output was treated as a correct R00 category-level mapping (abnormalities of heartbeat). Future work should improve subcode-level specificity and clinical precision in *ICD-10-CM* assignment.

The red-flag keyword filter used during postprocessing was intentionally conservative and improved overall performance, but it was also a coarse rule-based step. For example, because the keyword list included terms such as “explicitly,” valid outputs could theoretically be removed if the model echoed instruction-like language in its structured response. Although postprocessing improved overall F_1-scores, this approach may have increased precision at the cost of removing some true positive outputs and should be refined in future work.

In addition, the scope of this study was restricted to cardiorespiratory S&S, which limits generalizability to other clinical domains. Future work should expand the dataset, incorporate a broader range of symptom domains, and further refine prompt engineering and multimodule orchestration strategies to improve robustness and adaptability across diverse biomedical contexts.

This study did not systematically evaluate model-specific runtime options, such as reasoning effort levels or structured output enforcement, because the primary objective was to assess the effects of prompt design and workflow structure under a fixed inference setting. We also did not compare our framework against proprietary frontier models, such as GPT-4. Although such models could provide an additional reference benchmark, they would not represent a fixed upper performance bound because extraction accuracy depends on factors such as model version, prompting strategy, runtime configuration, and evaluation design. In addition, this study was intentionally focused on privacy-preserving, locally deployable workflows for clinical text processing. Future work should investigate whether runtime configuration options further improve output stability and extraction accuracy and should include direct comparisons between open-weight local models and high-performing closed commercial models under standardized prompts and evaluation settings.

Conclusions

In this study, Llama 3.3-70B was used primarily for prompt development to support the extraction of clinical S&S and their mapping to *ICD-10-CM* codes. We then evaluated the final multimodule LLM framework with postprocessing on an independent test set of 500 clinical notes using both Llama 3.3-70B and gpt-oss-120B. Under the entity-level microaveraged evaluation across the full corpus, Llama 3.3-70B achieved F_1-scores of 0.73 for S&S extraction and 0.69 for *ICD-10-CM* code generation, whereas gpt-oss-120B achieved F_1-scores of 0.88 and 0.87, respectively. In note-level macroaveraged evaluation, performance was higher, with Llama 3.3-70B achieving F_1-scores of 0.95 for both S&S extraction and *ICD-10-CM* mapping and gpt-oss-120B achieving F_1-scores of 0.97 and 0.98. Overall, the framework demonstrates the feasibility of locally deployable LLMs for structured extraction of cardiorespiratory S&S from clinical notes and may increase the level of privacy safety by allowing on-premises processing without external data transmission.

XSL•FO
RenderX

Acknowledgments

The authors used the generative artificial intelligence tool Llama 3.3-70B by Meta and gpt-oss by OpenAI to extract symptoms and signs from clinical notes. ChatGPT by OpenAI was used for grammar correction during manuscript generation.

Funding

The project was funded in part by National Institutes of Health grant R33HL143317.

Data Availability

The datasets analyzed during this study are available in the MTSamples database [16]. The labeled data in our study are available from the corresponding author on reasonable request. The code can be accessed at [27].

Authors' Contributions

Conceptualization: YB, JF
Data curation: YB, WC, JF
Formal analysis: YB
Funding acquisition: JF
Investigation: YB
Methodology: YB
Resources: WC, JF
Software: YB
Supervision: JF
Validation: YB, WC, JF
Visualization: YB
Writing – original draft: YB
Writing – review & editing: WC, JF

Conflicts of Interest

None declared.

References

1. Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. Thorax 2005 Dec;60(12):1059-1065. [doi: 10.1136/thx.2005.045880] [Medline: 16227326]
2. Prado MG, Kessler LG, Au MA, et al. Symptoms and signs of lung cancer prior to diagnosis: case-control study using electronic health records from ambulatory care within a large US-based tertiary care centre. BMJ Open 2023 Apr 20;13(4):e068832. [doi: 10.1136/bmjopen-2022-068832] [Medline: 37080616]
3. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013 May;143(5 Suppl):e142S-e165S. [doi: 10.1378/chest.12-2353] [Medline: 23649436]
4. Choi E, Bahadori MT, Kulas JA, Schuetz A, Stewart WF, Sun J. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. Presented at: Advances in Neural Information Processing Systems (NeurIPS); Dec 5-10, 2016 URL: https://papers.neurips.cc/paper/2016/file/231141b34c82aa95e48810a9d1b33a79-Paper.pdf [accessed 2026-05-07]
5. Shah-Mohammadi F, Finkelstein J. NLP-assisted differential diagnosis of chronic obstructive pulmonary disease exacerbation. Stud Health Technol Inform 2024 Jan 25;310:589-593. [doi: 10.3233/SHTI231033] [Medline: 38269877]
6. Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. arXiv. Preprint posted online on Apr 10, 2019. [doi: 10.48550/arXiv.1904.05342]
7. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021 May 20;4(1):86. [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]
8. Li Y, Rao S, Solares JRA, et al. BEHRT: transformer for electronic health records. Sci Rep 2020;10(1):7155. [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
9. Choi HS, Song JY, Shin KH, Chang JH, Jang BS. Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. Radiat Oncol J 2023 Sep;41(3):209-216. [doi: 10.3857/roj.2023.00633] [Medline: 37793630]

XSL•FO
RenderX

10. Shah-Mohammadi F, Finkelstein J. Extraction of substance use information from clinical notes: generative pretrained transformer-based investigation. JMIR Med Inform 2024 Aug 19;12:e56243. [doi: 10.2196/56243] [Medline: 39037700]
11. Li Y, Mamouei M, Salimi-Khorshidi G, et al. Hi-BEHRT: hierarchical transformer-based model for accurate prediction of clinical events using multimodal longitudinal electronic health records. IEEE J Biomed Health Inform 2023;27(2):1106-1117. [doi: 10.1109/JBHI.2022.3224727] [Medline: 36427286]
12. Wang C, Long Q, Xiao M, et al. BioRAG: a RAG-LLM framework for biological question reasoning. arXiv. Preprint posted online on August 2, 2024. [doi: 10.48550/arXiv.2408.01107]
13. Shah-Mohammadi F, Finkelstein J. Utilizing RAG and GPT-4 for extraction of substance use information from clinical notes. Stud Health Technol Inform 2024 Nov 22;321:94-98. [doi: 10.3233/SHTI241070] [Medline: 39575787]
14. Bai Y, Cui W, Finkelstein J. Performance of open-source large language models to extract symptoms from clinical notes. Stud Health Technol Inform 2025 Aug 7;329:663-667. [doi: 10.3233/SHTI250923] [Medline: 40775941]
15. Bai Y, Cui W, Finkelstein J. Utilizing open-source large language models to extract genitourinary symptoms from clinical notes. Stud Health Technol Inform 2025 Jun 26;328:16-20. [doi: 10.3233/SHTI250664] [Medline: 40588872]
16. MTSamples.com. URL: https://www.mtsamples.com/ [accessed 2026-05-11]
17. Symptoms and signs involving the circulatory and respiratory systems R00-R09. ICD10Data. URL: https://www.icd10data.com/ICD10CM/Codes/R00-R99/R00-R09 [accessed 2026-05-11]
18. Llama3.3. Ollama. 2024. URL: https://ollama.com/library/llama3.3 [accessed 2026-05-11]
19. Agarwal S, Ahmad L, Ai J, Altman S, Applebaum A, Arbus E, et al. Gpt-oss-120b & gpt-oss-20b model card. arXiv. Preprint posted online on Aug 8, 2025. [doi: 10.48550/arXiv.2508.10925]
20. Wu Q, Bansal G, Zhang J, et al. AutoGen: enabling next-gen LLM applications via multi-agent conversation. arXiv. Preprint posted online on Aug 16, 2023. [doi: 10.48550/arXiv.2308.08155]
21. Li G, Hammoud HAK, Itani H, Khizbullin D, Ghanem B. CAMEL: communicative agents for “mind” exploration of large language model society. arXiv. Preprint posted online on Mar 31, 2023. [doi: 10.48550/arXiv.2303.17760]
22. Hong S, Zhuge M, Chen J, Zheng X, Cheng Y, Wang J, et al. MetaGPT: meta programming for a multi-agent collaborative framework. Presented at: The Twelfth International Conference on Learning Representations (ICLR 2024); May 7-11, 2024 URL: https://openreview.net/forum?id=VtmBAGCN7o [accessed 2026-05-07]
23. Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, et al. Large language model based multi-agents: a survey of progress and challenges. Presented at: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24); Aug 3-9, 2024; Jeju Island, South Korea p. 8048-8057. [doi: 10.24963/ijcai.2024/890]
24. Smiley A, Reategui-Rivera CM, Villarreal-Zegarra D, Escobar-Agreda S, Finkelstein J. Exploring artificial intelligence biases in predictive models for cancer diagnosis. Cancers (Basel) 2025 Jan 26;17(3):407. [doi: 10.3390/cancers17030407] [Medline: 39941776]
25. Finkelstein J, Gabriel A, Schmer S, Truong TT, Dunn A. Identifying facilitators and barriers to implementation of AI-assisted clinical decision support in an electronic health record system. J Med Syst 2024 Sep 18;48(1):89. [doi: 10.1007/s10916-024-02104-9] [Medline: 39292314]
26. Kawamoto K, Finkelstein J, Del Fiol G. Implementing machine learning in the electronic health record: checklist of essential considerations. Mayo Clin Proc 2023 Mar;98(3):366-369. [doi: 10.1016/j.mayocp.2023.01.013] [Medline: 36868743]
27. Resources. Arizona Telemedicine Program. URL: https://telemedicine.arizona.edu/resources [accessed 2026-05-08]

Abbreviations

COPD: chronic obstructive pulmonary disease
EM: extraction module
***ICD-10-CM*:** *International Classification of Diseases, Tenth Revision, Clinical Modification*
LLM: large language model
NLP: natural language processing
RM: refinement module
S&S: symptom and sign

Edited by A Coristine; submitted 11.Dec.2025; peer-reviewed by Á García-Barragán, M Torii; revised version received 28.Apr.2026; accepted 28.Apr.2026; published 22.May.2026.

Please cite as:
Bai Y, Cui W, Finkelstein J
Extracting Cardiorespiratory Symptoms From Clinical Notes Using Open-Weight Large Language Models: Method Development and Validation Study
JMIR Cardio 2026;10:e89480
URL: https://cardio.jmir.org/2026/1/e89480
doi:10.2196/89480

XSL•FO
RenderX

© Yunbing Bai, Wanting Cui, Joseph Finkelstein. Originally published in JMIR Cardio (https://cardio.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Heart Rate Estimation Using the Galaxy Watch During Maximal Cardiopulmonary Exercise Testing: Cross-Sectional Validation Study

Allan Inoue, PhD; João Paulo Ferreira Soares, MSc; Felipe Antunes-Santos, MSc; Alexandre Ferreira, MSc; Alberto Gonçalves, BSc; João Arthur Alcântara, PhD; Marcelo Rodrigues dos Santos, PhD

Health Data Lab, Sidia Institute of Science and Technology, Av. Autaz Mirim, 2211 - Distrito Industrial I, Manaus, Brazil

Corresponding Author:
Marcelo Rodrigues dos Santos, PhD
Health Data Lab, Sidia Institute of Science and Technology, Av. Autaz Mirim, 2211 - Distrito Industrial I, Manaus, Brazil

Abstract

Background: Photoplethysmography-based smartwatches are increasingly used for continuous heart rate (HR) monitoring. Their accuracy has been demonstrated at rest or during low-intensity activity, but data are scarce for maximal-intensity exercise, when motion artifacts and rapid hemodynamic changes can degrade the photoplethysmography signal. Validating these devices under such demanding conditions is essential before they are applied to clinical exercise testing, athletic training, or remote health monitoring.

Objective: This study aimed to evaluate the validity of the Samsung Galaxy Watch 6 (GW6) in estimating HR throughout a graded, maximal ramp cardiopulmonary exercise test performed on a treadmill. A secondary aim was to explore whether measurement error varies across 5 predefined intensity zones (50% - 60%, 60% - 70%, 70% - 80%, 80% - 90%, and 90% - 100% of the maximum HR determined individually for each participant).

Methods: Overall, 55 healthy adults (30 men, 25 women; mean age 30.3, SD 8.2 years) completed a symptom-limited incremental treadmill protocol to volitional exhaustion. Simultaneous HR recordings were obtained from the GW6 (left arm) and a Polar H10 chest strap monitor, which served as the reference standards. For each intensity zone, the following agreement indices were computed: intraclass correlation coefficient (ICC), median absolute error, median absolute percentage error, and root mean squared error. Bland-Altman analysis was performed to quantify the mean bias and 95% limits of agreement between the GW6 and the Polar H10. Statistical significance was set at $P<.05$.

Results: Agreement between the GW6 and Polar H10 varied across exercise intensities. ICC indicated moderate to good agreement at low to moderate intensities (ICC=0.71 at 50% - 60%; ICC=0.89 at 60% - 70%; ICC=0.54 at 70% - 80%; and ICC=0.64 at 80% - 90% HRmax), and at 90% - 100% of HRmax the agreement was good-to-excellent (ICC=0.90). Absolute error metrics showed stable or reduced errors with increasing intensity, with median absolute error consistently around 1 - 3 bpm and median absolute percentage error declining from 2.90% at 50% - 60% HRmax to 0.60% - 0.75% at ≥70% HRmax. Root mean squared error ranged from 4.62 to 4.88 bpm across intensity zones. Bland-Altman analysis showed that the GW6 consistently underestimated HR compared with the Polar H10, with an overall mean bias of −2.67 bpm and wide limits of agreement (−16.90 to 11.57 bpm). This negative bias was present across all HR zones. The agreement was adequate for group-level comparisons but displayed substantial individual variability.

Conclusions: The GW6 provides a good degree of validity for HR monitoring during a maximal treadmill cardiopulmonary exercise test in healthy young adults. Although measurement error increases modestly at near-maximal workloads, absolute errors remain well within clinically acceptable thresholds. These findings support the potential use of GW6 as a convenient, noninvasive alternative for HR tracking in laboratory-based exercise testing.

(JMIR Cardio 2026;10:e81917) doi:10.2196/81917

KEYWORDS

photoplethysmography; smartwatches; wearable; maximal exercise; cardiopulmonary exercise test

Introduction

The use of wrist-worn wearables (eg, smartwatches) has been growing rapidly over the last years. These devices have become popular for monitoring various health parameters, including heart rate (HR) through photoplethysmography sensors. HR is estimated from photoplethysmography signals through the changes in blood volume beneath the skin [1]. Although convenient and noninvasive, photoplethysmography-based HR monitoring remains vulnerable to motion artifacts, sensor-skin

XSL•FO
RenderX

interface issues, and light interference, particularly during dynamic or high-intensity exercise. Ensuring the validity of HR data under such demanding conditions is therefore critical, especially when wearables are used for exercise prescription or physiological assessment [2].

Several studies have evaluated the accuracy of wrist-based devices at rest or during light-to-moderate exercise, showing acceptable validity compared to electrocardiogram (ECG) or chest-strap references [3]. Thomson et al [3] compared HR values from the Apple Watch and Fitbit Charge HR2 with ECG during the Bruce Protocol across multiple intensity domains, from very light to very vigorous exercise. The Apple Watch showed lower relative error rates (2.4% - 5.1%) than the Fitbit (3.9% - 13.5%) across all intensities, with very high concordance (concordance correlation coefficient >0.90) at light workloads. However, HR accuracy progressively declined at higher intensities, indicating that optical sensors may lose precision during vigorous or near-maximal exertion. Importantly, that study did not include a maximal cardiopulmonary exercise test (CPET) or direct measurement of oxygen consumption.

To date, only 1 study [4] (conducted with the Apple Watch during a maximal CPET using a ramp protocol and simultaneous gas exchange analysis) has validated HR estimation under maximal physiological stress. The authors reported good to very good correlations with criterion HRmax and trivial mean bias compared with a Polar chest strap, although moderate interdevice variability was observed. These findings demonstrate that wrist-worn optical sensors can validly estimate HRmax during maximal effort, but evidence remains limited to a single device and a small sample size.

Given the lack of studies assessing the accuracy of smartwatches during maximal effort protocols, particularly those incorporating direct measurement of oxygen consumption, further validation is warranted. To date, no study has systematically evaluated the HR performance of the Samsung Galaxy Watch 6 (GW6) in young healthy individuals during a ramp incremental test with concurrent oxygen uptake analysis. By integrating continuous gas-exchange assessment with a standardized graded treadmill protocol, this study provides the first validation of the GW6 under conditions that reflect both physiological demand and methodological rigor; the only available investigation involved patients with coronary artery disease [5]. Accurate HR monitoring during CPET is fundamental for determining metabolic thresholds, exercise intensity zones, and overall cardiorespiratory fitness.

Therefore, this study aimed to determine the validity of the Samsung GW6 in estimating HR during a maximal ramp CPET performed on a treadmill in healthy adults, using the Polar H10 chest strap as the criterion measure. We hypothesized that HR values obtained from the GW6 would demonstrate at least moderate agreement (intraclass correlation coefficient [ICC] ≥0.60) with the reference device across 5 intensity zones (Z1–Z5) of the maximal ramp protocol.

Methods

Study Design

This is a cross-sectional study, within - subject design to compare the HR accuracy from the GW6 with criterion measures of HR (Polar H10 chest strap) during a maximal ramp CPET conducted on a treadmill. All participants had their health conditions evaluated by a physician before the tests. Participants were instructed to abstain from alcoholic beverages and any medication for 24 hours, to avoid caffeine intake for 12 hours before the test, and to hydrate with water only. Body weight and height were measured using a weighing scale and stadiometer (InBody 770 analyzer), and skin type was measured via the Fitzpatrick skin type scale [6]. The environment in the laboratory was standardized throughout all sessions by maintaining an ambient temperature of approximately 22°C. All technical procedures followed the American College of Sports Medicine guidelines [7]. We followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement: guidelines for reporting observational studies [8].

Participants

The inclusion criteria were healthy, recreationally active participants who engaged in at least 150 minutes of moderate-intensity activities and/or 75 minutes of vigorous activities per week.

Exclusion criteria were overweight and obesity, recent musculoskeletal injuries, any cardiovascular disease, and use of medications or supplements (eg, beta-blockers and caffeine) that impact HR measures. Participants signed the informed consent form after a verbal and written explanation of the activities to be carried out and a full understanding of the possible risks involved in participating in the study.

Initially, 76 participants began the CPET protocol. Four were excluded because of execution failures, and 17 were excluded because of device failures. The participant selection flowchart can be found in the Results section. Execution failure refers to problems that occurred before the trial started. Examples include the device not being set up correctly; failure to pair or synchronize the device; and incorrect entry of participant information into the device. Device failure refers to problems that arose during or after the trial. Examples include: loss of connection during CPET execution; the Polar H10 chest strap producing an unstable, missing, or abruptly dropping HR signal for any portion of the test; and devices that were correctly configured but still failed to generate all required output files.

Maximal Ramp Cardiopulmonary Exercise Test

All participants performed maximal ramp CPET on a motorized treadmill (COSMED T170 DE) to determine HR at increasing intensities and maximal oxygen consumption (determined as the highest value observed during the test). The test was started at 5 km·h^-1 and the speed increased by 1 km·h^-1 every minute, with the slope kept constant at 1% [9] until they met at least two out of the following three conditions: (1) reached 90% of their maximal HR value calculated using Tanaka's equation [10], (2) achieved a respiratory exchange ratio (RER) greater

XSL•FO
RenderX

than or equal to 1.05 [11], and (3) reached maximal voluntary exhaustion, evaluated by the Borg (6-20) subjective scale of effort [12]. Gas exchange recording was performed using the K5 portable metabolic analyzer (COSMED). Device setup, calibration, and data extraction were conducted via COSMED OMNIA Metabolic software. The equipment was calibrated before each test as recommended by the manufacturer.

The test was terminated according to the termination criteria recommended by the American College of Sports Medicine guidelines [13], which include the occurrence of cardiovascular, respiratory, or musculoskeletal symptoms, abnormal blood pressure responses, significant rhythm disturbances, or voluntary exhaustion. All tests were supervised by a physician trained in CPET.

During the test, participants wore a chest strap HR monitor (Polar H10) and the GW6 smartwatch. HR data were evaluated across 5 intensity zones commonly used in various applications (50% - 60%, 60% - 70%, 70% - 80%, 80% - 90%, and 90% - 100% of HRmax). The average HR was calculated for each of the 5 intensity zones based on the maximal HR measured by the Polar H10 from the whole test. This approach is consistent with standard clinical practice in cardiology and incremental exercise testing and was chosen to provide a stable and representative estimate of the physiological response at each workload while minimizing the influence of beat-to-beat variability and transient signal artifacts.

Resting HR was measured using the Polar device before the start of the $\dot{V}O_2$ max test protocol. Participants were seated quietly for several minutes to ensure a stable baseline prior to data collection.

Criterion Measure

The Polar H10 chest strap (Polar Electro Oy) served as the criterion method for measuring HR. This device has previously been validated against ECG [14]. Each participant was fitted with the Polar chest strap, set to collect continuous HR data throughout the protocol. On completion of the activity, data were transferred from each Polar transmitter to a central laptop to be viewed and analyzed using Polar Flow (version 6.25.2, Polar Electro Oy).

Wrist-Worn Wearable Device

The GW6 (Samsung Electronics Co Ltd) is a multisensor smartwatch. Accelerometer, gyroscope, barometer, magnetometer, and photoplethysmography are some of the GW6's sensors. During our experiments, all GW6s were connected to separate smartphones (Galaxy A22, Samsung Electronics Co Ltd), and all HR data were exported using the Samsung Health app. In cases where the activity data needed to be trimmed, full health data were exported via the Samsung Health app, with the raw data CSV file processed using Python (version 3.12.7; Python Software Foundation).

Participant height, date of birth, and wrist orientation were entered into each device before the protocol. The GW6 was positioned on the left arm located just above the wrist, as per manufacturer instructions (2 cm above styloid process of radius). In addition, the GW6 was set to record the activity protocol as a workout, being started and stopped at the commencement and completion of the activity. The GW6 was set to record "treadmill run," as this was the most relevant option available on the watches.

Statistical Analysis

Sample size calculation was performed on a web-based calculator according to Arifin [15]. The lowest acceptable ICC was 0.50, and the expected reliability was ICC=0.75 [16]. The number of raters or repetitions considered was 2 (Polar H10 vs GW6). A significance level of α=.05 (2-tailed) and a power of 80% were set. As a result, 45 participants were needed for this study. However, considering a dropout rate of ~15%, we recruited 55 healthy adults, composed of 30 men and 25 women, aged between 18 and 59 years.

All data were analyzed using a custom-made Microsoft Excel spreadsheet [17]. Data are shown as mean (SD) or median (IQR). The validity of the wearable was determined by the following statistical tests: (1) Correlation between the wearable and the criterion measurement: ICC, common cut-off points for validity assessment: poor (<0.50), moderate (0.50 to 0.75), good (0.75 to 0.90) and excellent (>0.90) [16]; (2) Pearson correlation coefficient; (3) Systematic differences between the wearable and the criterion measurement: median absolute error (MAE) and median absolute percentage error (MAPE) compared to the criterion measurement (criterion-wearable)/criterion × 100). The MAPE was considered with acceptable error when <10% [18]. Additionally, root mean square error (RMSE) was calculated. A predefined subgroup analysis was conducted by stratifying participants according to skin tone using the Fitzpatrick classification. To preserve statistical power, Fitzpatrick categories were aggregated into 3 groups based on pigmentation level. Device performance was then evaluated within each group using the same accuracy and agreement metrics across exercise intensity zones. This subanalysis aimed to assess whether skin tone influenced HR estimation during the incremental cardiopulmonary exercise test.

Bland-Altman analysis was performed to quantify the mean bias and 95% limits of agreement (LoA) between the GW6 and the Polar H10 across the full dataset and within each predefined HR zone (50% - 60%, 60% - 70%, 70% - 80%, 80% - 90%, and 90% - 100% HRmax). For each comparison, the mean difference (device − reference), the corresponding SD, and the upper and lower LoA (mean difference, 1.96 SD) were calculated. Visual inspection was used to assess whether the differences were homoscedastic and whether any proportional bias was present across the range of HR values.

To further characterize agreement between the GW6 and the Polar H10 reference, we computed the coefficient of variation (CV%) for each intensity zone as CV%=(SD/mean) × 100, applying the formula separately to the GW6 and H10 HR values. Additionally, 2-way repeated measures ANOVA was performed to examine whether HR values obtained from GW6 differed from those recorded with the Polar H10 across the 5 predefined intensity zones (% HRmax).

XSL•FO
RenderX

Ethical Considerations

The study protocol was approved by the Investiga Instituto de Pesquisa Ethics Committee (CAAE 62031722.4.0000.5599), and all procedures were conducted following the standards set by the Declaration of Helsinki. Data were anonymized before analysis to ensure confidentiality. No monetary or other forms of compensation were provided. The manuscript contains no identifiable information or images of individual participants. All participants were informed about the study procedures and provided written informed consent.

Results

Participant characteristics are presented in Table 1 and the flow diagram of participants is presented in Figure 1. The incremental CPET was stopped only after participants satisfied at least 2 of 3 predefined maximal-effort criteria: a Borg rating of perceived exertion ≥18, indicating volitional exhaustion; a HR reaching ≥90% of the age-predicted maximal HR calculated with Tanaka's equation; and RER ≥1.05, reflecting near-maximal metabolic stress. Of the 55 participants, all 55 (100%) met the HR criterion, 31 (56%) achieved a Borg score ≥18, and 40 (73%) attained an RER ≥1.05 (mean RER at termination 1.08, SD 0.07). When the criteria were examined in combination, 31 participants fulfilled both the Borg and HR conditions, 16 participants met both the Borg and RER conditions, and 40 participants satisfied the HR and RER conditions; among these, 16 participants fulfilled all 3 criteria simultaneously.

Table . Participant characteristics.

Variables	Male (n=30)	Female (n=25)	All (n=55)
Age (years), mean (SD)	29.8 (8.6)	31 (7.7)	30.3 (8.2)
BMI (kg/m^2), mean (SD)	71.7 (8.1)	60 (9.6)	66.3 (10.6)
Height (cm), mean (SD)	170 (6.1)	158.5 (4.7)	164.8 (8)
Fitzpatrick skin type, n (%)			
II	0 (0)	1 (1.8)	1 (1.8)
III	4 (7.3)	10 (18.2)	14 (25.5)
IV	16 (29.1)	12 (21.8)	28 (50.9)
V	8 (14.5)	2 (3.6)	10 (18.2)
VI	2 (3.6)	0 (0)	2 (3.6)
Resting HR[a] (bpm[b]), mean (SD)	63.1 (8.7)	70.8 (9.7)	66.6 (9.9)
RER[c]max[d], mean (SD)	1.09 (0.04)	1.06 (0.08)	1.08 (0.07)
$\dot{V}O_2$ max[e] (mL·kg·min-1), mean (SD)	46.8 (5.5)	41.4 (7.7)	44.3 (7.1)

[a]HR: heart rate.

[b]bpm: beats per minute.

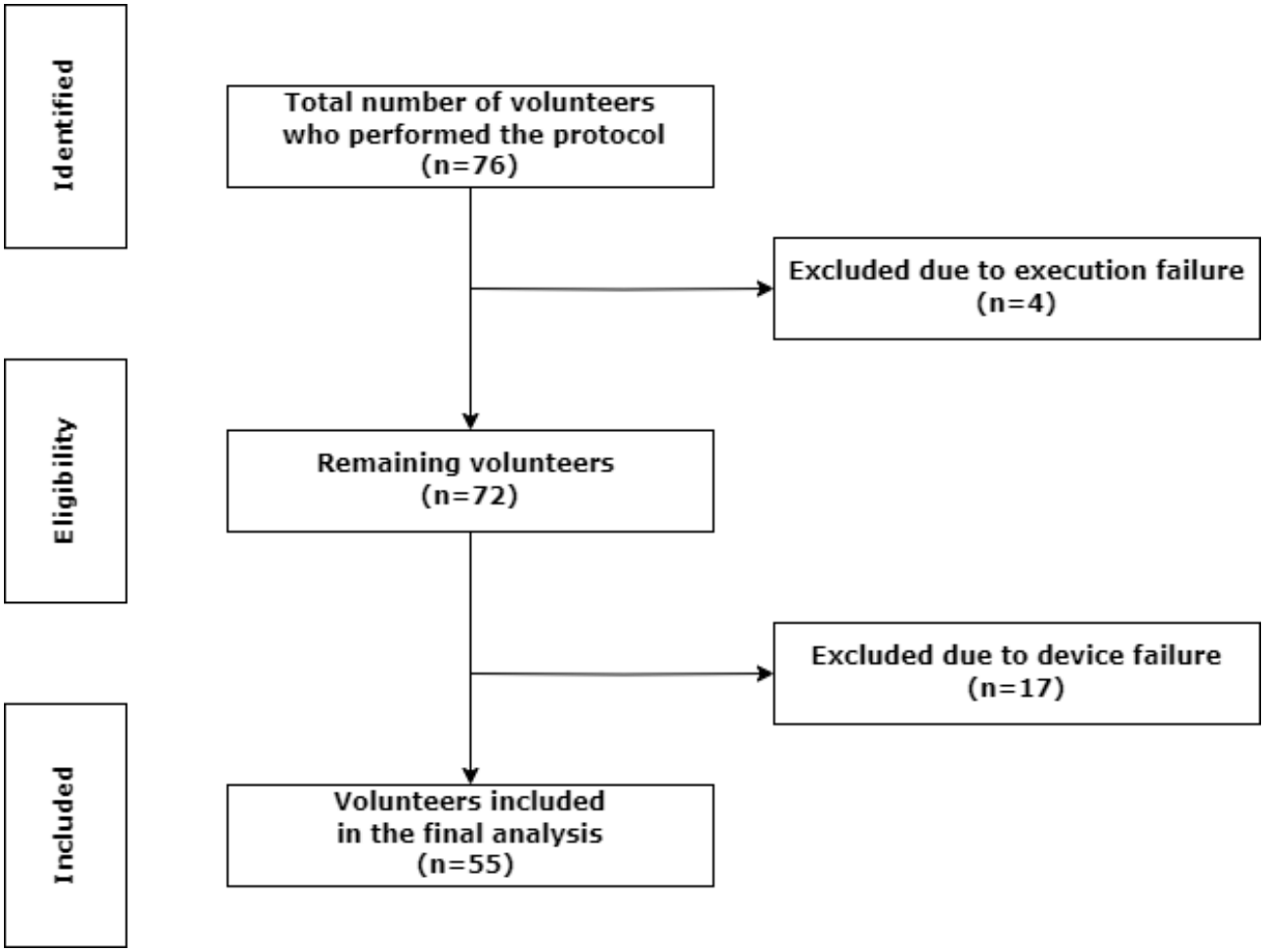
[c]RER: respiratory exchange ratio.

[d]max: maximum.

[e]$\dot{V}O_2$ max: maximal oxygen consumption.

XSL•FO
RenderX

Figure 1. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) flow diagram.



Descriptive values, correlations (Pearson and ICC), MAE, MAPE, and RMSE are presented in Table 2. At low intensities (50% - 60% HRmax), agreement was moderate (ICC=0.71; 95% CI 0.55-0.82), with a MAE of 3.0 bpm and a MAPE of 2.9%. Agreement improved at moderate intensities (60% - 70% HRmax), reaching good levels (ICC=0.89; 95% CI 0.82-0.93), accompanied by a reduction in relative error (MAPE=1.77%). At vigorous intensities, ICC values declined, indicating reduced between-subject agreement (ICC 0.54; 95% CI 0.32-0.70 at 70% - 80% HRmax and ICC=0.64; 95% CI 0.45-0.77 at 80% - 90% HRmax). Despite this reduction, absolute and relative error metrics improved substantially, with MAE approximately 1.0 bpm and MAPE below 1%, suggesting higher point-wise accuracy. At maximal intensities (90% - 100% HRmax), agreement increased markedly, reaching good-to-excellent levels (ICC=0.90; 95% CI 0.83-0.94), while error metrics remained low (MAE=1.0 bpm; MAPE=0.60%). RMSE values were relatively consistent across intensity zones, ranging from 4.37 to 4.88 bpm.

XSL•FO
RenderX

Table . Galaxy Watch 6 heart rate and comparisons with criterion (Polar H10). Median absolute error and median absolute percentage error are reported as median (IQR). In the FITZ column: Groups II-III: 15 participants; IV: 28 participants; V-VI: 12 participants; all: 55 participants.

%HRmax[a] and FITZ[b]		H10[c] (BPM[d]), mean (SD)	GW6[e] (BPM)	ICC[f] (CI 95%)	Pearson *r*[g] (CI 95%)	Median AE[h] (IQR)	Median APE[i] (IQR)	RMSE[j]
50-60								
	II-III	99.87 (6.01)	95.40 (4.82)	0.73 (0.36-0.90)	0.73 (0.35-0.90)	3 (2-4)	3.16 (2.06-3.93)	4.57
	IV	99.04 (4.92)	95 (5.52)	0.62 (0.33-0.81)	0.62 (0.32-0.81)	3 (2-4)	2.97 (1.98-4.07)	5.23
	V-VI	97.17 (5.57)	95.42 (6.11)	0.93 (0.77-0.98)	0.92 (0.74-0.98)	1 (0-1)	1.02 (0-1.06)	2.26
	All	98.85 (5.36)	95.20 (5.38)	0.71 (0.55-0.82)	0.69 (0.53-0.81)	3 (1-4)	2.90 (1.04-3.99)	4.62
60-70								
	II-III	118.07 (5.95)	113.53 (5.96)	0.68 (0.27-0.88)	0.66 (0.22-0.88)	3.50 (1.25-5)	3.10 (1.08-4.21)	4.64
	IV	116.89 (5.12)	113.25 (5.89)	0.58 (0.27-0.78)	0.58 (0.26-0.78)	2 (1-5)	1.79 (0.86-4.26)	4.99
	V-VI	115.67 (5.16)	114.42 (5.95)	0.97 (0.90-0.99)	0.97 (0.91-0.99)	1 (0-1.50)	0.83 (0-1.34)	1.35
	All	116.95 (5.33)	113.58 (5.83)	0.89 (0.82-0.93)	0.90 (0.83-0.94)	2 (1-5)	1.77 (0.85-4.05)	4.37
70-80								
	II-III	136.27 (7.24)	132.87 (6.82)	0.81 (0.52-0.93)	0.80 (0.48-0.93)	1.50 (1-2.75)	1.06 (0.73-2.01)	3.28
	IV	135.61 (5.78)	134.45 (10.46)	0.42 (0.06-0.68)	0.49 (0.14-0.73)	1 (1-2.75)	0.75 (0.74-2.03)	6.05
	V-VI	133.83 (6.75)	133.67 (6.87)	0.99 (0.96-1)	0.99 (0.96-1)	0 (0-1)	0 (0-0.75)	0.67
	All	135.40 (6.35)	133.85 (8.77)	0.54 (0.32-0.70)	0.70 (0.53-0.81)	1 (1-2)	0.75 (0.69-1.52)	3.96
80-90								
	II-III	155 (7.73)	151.87 (7.55)	0.46 (–0.04 to 0.78)	0.45 (–0.08 to 0.78)	1 (0-3)	0.66 (0-1.82)	2.83
	IV	154.50 (6.79)	153.21 (11.92)	0.39 (0.02-0.66)	0.44 (0.08-0.70)	1 (0-2)	0.64 (0.00 to 1.30)	7.71
	V-VI	151.58 (7.74)	150.92 (7.74)	0.96 (0.88-0.99)	0.96 (0.86-0.99)	0 (0-1)	0 (0-0.68)	1.17
	All	154 (7.24)	152.35 (9.95)	0.64 (0.45-0.77)	0.75 (0.60-0.84)	1 (0-2)	0.64 (0-1.34)	4.88
90-100								
	II-III	174.67 (8.81)	169.53 (10.93)	0.53 (0.05-0.81)	0.53 (0.02-0.82)	1 (0.25-3)	0.61 (0.14-1.62)	6.23
	IV	173.61 (7.75)	170.89 (8.97)	0.77 (0.57-0.89)	0.78 (0.57-0.89)	1 (0-3.50)	0.58 (0-2.01)	3.23
	V-VI	170.67 (6.97)	169.17 (5.57)	0.79 (0.41-0.93)	0.79 (0.39-0.94)	1 (1-1)	0.60 (0.58-0.62)	1.38
	All	173.25 (7.88)	170.15 (8.84)	0.90 (0.83-0.94)	0.87 (0.79-0.92)	1 (0.75-3)	0.60 (0.41-1.62)	4.52

[a]%HRmax: percentage of maximum heart rate.

[b]FITZ: Fitzpatrick Skin group.

[c]H10: Polar H10.

[d]BPM: beats per minute.

XSL•FO
RenderX

[e]GW6: Galaxy Watch 6.

[f]ICC: intraclass correlation coefficient.

[g]Pearson *r*: Pearson correlation coefficient.

[h]AE: absolute error.

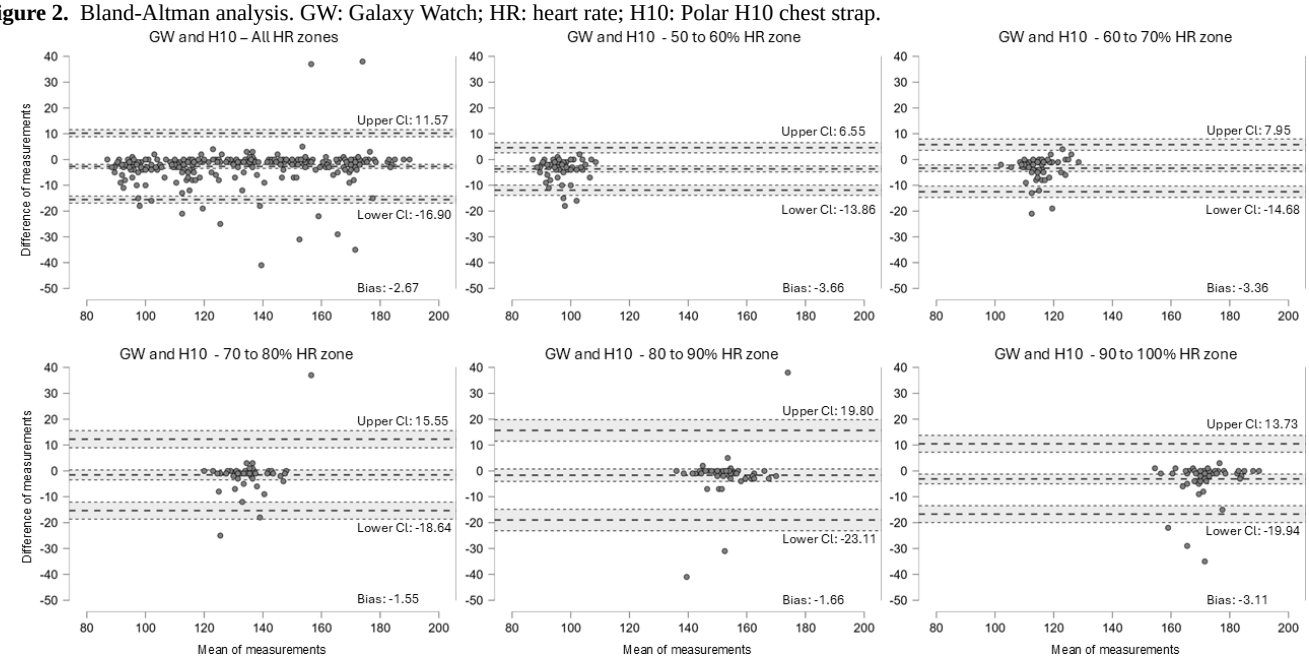
[i]APE: absolute percentage error.

[j]RMSE: root mean square error.

Across all HR measurements, the GW6 showed a systematic underestimation of HR relative to the Polar H10, with a mean bias of −2.67 bpm and 95% LoA ranging from −16.90 to 11.57 bpm. Although most points fell within these limits, the negative bias indicates that the smartwatch consistently reported lower values than the reference. When examined by HR zones, the pattern of underestimation persisted with small variations in magnitude: 50% - 60% HRmax: bias −3.66 bpm, LoA −13.86 to 6.55 bpm; 60% - 70% HRmax: bias −3.36 bpm, LoA −14.68 to 7.95 bpm; 70% - 80% HRmax: bias −1.55 bpm, LoA −18.64 to 15.55 bpm; 80% - 90% HRmax: bias −1.66 bpm, LoA −23.11 to 19.80 bpm; and 90% - 100% HRmax: bias −3.11 bpm, LoA −19.94 to 13.73 bpm (Figure 2). Although agreement was acceptable for group-level analysis, the width of the LoA demonstrates meaningful variability at the individual level, particularly at higher intensities where physiological oscillations are greater.

Additionally, the CV ranged between 4.6% to 5.4% for Polar and 5.1% to 6.5% for GW. ANOVA (device) revealed a significant main effect of Device (P<.001), indicating that the GW6 records a modestly lower HR than the Polar H10. The main effect of % HR was also significant (P<.001), reflecting the expected increase in HR across intensity zones. Importantly, the interaction (device vs % HR) was not significant (P=.73), demonstrating that the systematic underestimation is consistent across all % HRmax levels.

Figure 2. Bland-Altman analysis. GW: Galaxy Watch; HR: heart rate; H10: Polar H10 chest strap.



Discussion

Principal Findings

The primary aim of this study was to evaluate whether the GW6 provides HR estimates that are in at least moderate agreement with a criterion reference chest-strap (Polar H10) across 5 intensity zones (Z1-Z5) of a maximal ramp CPET. The data partially support our hypothesis. Moderate agreement was observed at lower intensities (50% - 60% HRmax), improving to good agreement at moderate intensities (60% - 70% HRmax). At vigorous intensities (70% - 80% and 80% - 90% HRmax), ICC values declined, before increasing again at maximal effort (90% - 100% HRmax), where agreement reached good-to-excellent levels. These findings indicate that between-subject agreement is not monotonic across exercise intensity and is influenced by changes in HR variability across zones. Importantly, despite fluctuations in ICC, absolute and relative error metrics demonstrated a consistent improvement with increasing intensity. MAE remained low across all zones (approximately 1 - 3 bpm), while MAPE decreased progressively from 2.9% at low intensity to approximately 0.6% at vigorous-to-maximal intensities. These values are well within thresholds generally considered clinically acceptable for HR monitoring during exercise.

A “U-shaped” pattern in accuracy was observed, characterized by a decline in reliability (as measured by ICC) at intermediate workloads (70% - 80%), followed by a partial recovery at near-maximal intensities. This pattern reflects changes in the consistency of participant rankings (ICC) rather than absolute accuracy. While the ICC suggests reduced consistency in participant rankings at moderate to vigorous intensities, the absolute error metrics (MAE and MAPE) demonstrate a linear

XSL•FO
RenderX

improvement in accuracy as intensity increased. The decline in reliability at intermediate workloads is likely attributed to transient mechanical artifacts, such as increased variability and amplitude of arm motion, which may interfere with optical signal stability. Conversely, near-maximal effort, arm movement often becomes more rhythmically constrained, which may help stabilize the photoplethysmography signal and partially restore reliability. This pattern highlights the influence of workload-specific mechanical factors on the observed changes in reliability, rather than contradicting the general trend of increased physiological noise at higher intensities.

To date, to the best of our knowledge, only our study and the study of Abt et al [4] used CPET during HR measurement in healthy volunteers. When comparing our findings with those reported by Abt et al, both studies demonstrate good agreement between smartwatch-derived and criterion HR values, although the magnitude of validity differs according to exercise intensity and device model. Abt et al, with 15 participants, found very high correlations (r=0.87 - 0.98) and trivial bias for HRmax measurements obtained with the Apple Watch compared with a Polar reference, indicating strong validity at maximal effort. In contrast, our study, which included 55 participants, showed moderate to good reliability for the GW6 across exercise intensities, with the lowest agreement observed in the 80% - 90% HRmax range. Nevertheless, the MAPE remained small, supporting acceptable accuracy for practical use during exercise. Overall, the larger sample size of our study reinforces the consistency of these observations and suggests that, while optical sensors in current smartwatch models provide increasingly valid HR estimates, their performance may still vary across intensity domains and device types, particularly at submaximal workloads where motion artifacts are more likely.

In this study, the relative reliability (ICC) values obtained in the 5 HR zones are lower compared to those recently measured in individuals with coronary artery disease [5]. Kim et al [5] used an earlier-generation Samsung device (Galaxy Watch 4) and reported values above 0.90. The ICC values in this study ranged from moderate to excellent depending on exercise intensity, with the highest agreement observed at higher intensities (eg, 90 - 100% HR zone), while lower-to-moderate ICCs were observed in the remaining zones. These discrepancies can be partially explained by differences in study designs. Kim et al [5] evaluated individuals with coronary artery disease and used the modified Bruce protocol, while in this study, we selected healthy individuals and used a running exercise test protocol with 1-minute stages with constant inclination. In addition, the higher ICC observed by Kim et al [5] may be due to participants being instructed to hold a handrail in front of them as a safety measure during the CPET. As the hand position was fixed during the test, generating less movement artifacts, HR accuracy may have been overestimated. Additionally, with the aim of determining the accuracy of wearable devices (among them, the Samsung Gear S) to measure HR at rest (in the supine and seated positions) and during exercise (walking and running on a treadmill and cycling on a cycle ergometer) in healthy individuals, Wallen et al [19] reported ICC values for HR of 0.80 (0.40 to 0.93). Our results are similar if we consider the 95% CIs.

The MAE measurement of the Samsung Galaxy Watch 4, as reported by Lima et al [20] under hands-free treadmill conditions (2.92 bpm), aligns closely with our findings, which demonstrated a range of 1.00 to 3.00 bpm, depending on the intensity zone. In addition, a coefficient of variation ≤10% was arbitrarily suggested as the standard criterion used to establish an acceptable level of reliability [21]. We consider acceptable accuracy values to be between 5% - 10%.

Aiming to investigate the HR measurement accuracy of 3 wrist-worn devices (Apple Watch 6, Polar Vantage V, and Fitbit Sense) during various activities (sitting, walking, running, resistance exercises, and cycling), Hajj-Boutros et al [22] showed that Apple Watch 6 appears to have the highest level of accuracy for HR monitoring for all 5 activities, followed by Polar Vantage V and Fitbit Sense. However, these investigations were conducted using task-based protocols with fixed intensities, which differ substantially from the physiological and biomechanical demands of a continuous incremental CPET. This study extends this body of evidence by evaluating HR accuracy across predefined intensity zones during a maximal ramp CPET, a context characterized by progressively increasing cardiovascular load, changing movement dynamics, and greater susceptibility to motion-related artifacts. This approach provides complementary and clinically relevant information that is not captured by activity-based protocols, particularly regarding device performance under conditions of rapid HR transitions and high exercise intensities.

In contrast to our findings, Thomson et al [3] reported that MAPE increased with exercise intensity during a Bruce protocol test, which consisted of 3-minute stages with progressively increasing speed and incline until voluntary fatigue. Heart rate reserve (HRR) was categorized into 5 intensity zones: very light (<20% HRR), light (20% - 40% HRR), moderate (40% - 60% HRR), vigorous (60% - 85% HRR), and very vigorous (>85% HRR). Their results showed MAPE increasing from ~2.4% at the lowest intensity to ~4.7% at the highest [3]. In this study, we found a reduction in MAPE with increasing intensity. These discrepancies can be partially explained by the differences in maximal tests (Bruce vs 1-min stages and constant incline), smartwatch models (Apple Watch vs Galaxy Watch), and handrail use during testing (hands-on vs hands-free), and possibly differences in skin color type, which could influence photoplethysmography readings. Although our subgroup analyses stratified by Fitzpatrick skin type did not reveal a consistent pattern of worsening performance across intensity zones, some differences in ICC estimates were observed between skin tone groups in specific zones. However, these differences should be interpreted with caution, given the limited sample size within subgroups.

Our findings suggest that the GW6 can provide reasonably reliable HR data during a maximal treadmill test, which may be of interest to clinicians, exercise-science practitioners, and active individuals who seek a convenient alternative to laboratory-grade equipment. The device should be regarded as a supplemental tool rather than a replacement for ECG-based measurements when precise rhythm analysis or diagnostic accuracy is required. In settings where cost or accessibility limits the use of more expensive monitors, GW6 offers a simple,

XSL•FO
RenderX

affordable option for estimating HR, provided its limitations at near-maximal workloads are acknowledged. The use of stage-averaged HR in our study aligns with common practice in clinical exercise testing and CPET, facilitating physiological interpretation and comparability with previous studies. Although alternative approaches, such as shorter averaging windows or peak values within stages, may better capture rapid signal fluctuations, they may also be more susceptible to transient noise, particularly in wearable-derived measurements. Future studies specifically designed to compare different HR aggregation methods may help determine whether such alternatives provide additional value for device validation across different exercise intensities.

An operational aspect that emerged during data collection was the device failure rate of 22% (Figure 1), which reflects instances where the device did not generate usable HR data despite correct placement and adherence to protocol procedures. Although the focus of the study was accuracy during valid measurements, this level of attrition has practical implications. A failure rate of this magnitude may hinder consistent data acquisition in real-world laboratory or training environments, reducing the feasibility of using the device as a reliable continuous monitoring tool. This observation reinforces the need for device-specific robustness improvements and highlights the importance of considering both accuracy and operational reliability when evaluating wearable sensors.

Limitations

Some limitations of this study should be highlighted. The devices were assessed in young, healthy participants exercising in a standardized laboratory setting. Therefore, our study can only be generalized for this population and setting only. GW6 was placed in a standard position, not considering the effect of left or right side on the accuracy of HR measurement. Therefore, future studies could consider testing watch side randomization to assess HR accuracy. HR measurement was performed using the Polar H10 chest strap, which has a good correlation with ECG. However, this should be considered a limitation.

Every test was stopped only after 2 objective or subjective signs of near-maximal effort were present. For the participants who terminated the test after satisfying the HR and RER criteria without reaching Borg>18, the “90% - 100% intensity zone” may correspond to a submaximal range (eg, ≈81% - 90% of true physiological HRmax). This could slightly overestimate device accuracy because the devices were not challenged at the absolute physiological limits of those individuals.

This study evaluated the validity of GW6 only during a maximal ramp test performed on a treadmill. The continuous, relatively rhythmic arm swing that occurs while running may differ substantially from the movement patterns seen in other common training modalities such as resistance training, high-intensity interval training, or cycling, where wrist motion can be more erratic, static, or intermittent. Therefore, the performance of the embedded photoplethysmography algorithm observed in this investigation may not be directly applicable to those activities. Readers are cautioned against generalizing these findings beyond treadmill-based exercise, and future work should examine accuracy across a broader range of exercise modalities.

The study encountered a 22% device failure rate, leading to the exclusion of 17 participants due to unusable HR data. While the primary aim was to investigate accuracy during valid recordings, this attrition rate represents a limitation of the operational performance of the device. Such failures may impact the practical use in laboratory or field settings, and future research should investigate underlying causes and potential technical improvements to enhance data reliability. It should be noted that if device failure is correlated with specific physiological characteristics (eg, darker skin tone or specific arm movements), the remaining valid sample (n=55) may represent a “best-case” scenario, potentially inflating the accuracy estimates relative to the general population. However, the subgroup analyses stratified by skin tone according to the Fitzpatrick classification did not demonstrate statistically or clinically meaningful differences in device performance.

Additionally, we analyzed a potential survivability bias to identify any patterns among participants excluded. Initially, we conducted an empirical analysis of demographic variables between the failure group and the valid group, focusing on BMI, age, and skin pigmentation. No significant patterns were found in these variables. A second analysis was conducted to address concerns about failed trials from female participants, under the hypothesis that smaller wrist sizes might have affected the proper accommodation of smartwatches. However, this analysis also revealed no discernible patterns. While no bias was identified in the initial analysis, further investigations may be necessary.

Conclusions

This study indicates that the GW6 photoplethysmography-based technology demonstrated reasonable validity in monitoring HR during a maximal cardiopulmonary exercise test in healthy participants.

Acknowledgments

This work is the result of the RD&I project (VITALPPG), carried out by Sidia Institute of Science and Technology in partnership with Samsung Eletrônica da Amazônia Ltda, using resources from Federal Law number 8.387/1991, with its dissemination and advertising under article 39 of Decree number 10.521/2020. The partnership with the company refers solely to logistical and administrative support provided by the sponsor (eg, facilitating access to devices and coordinating timelines).

No generative AI tools were used in the generation of the manuscript text or data analysis. Only standard grammar was applied.

XSL•FO
RenderX

Funding

This study was funded by Samsung Eletrônica da Amazônia Ltda.

Data Availability

The raw data underlying the results reported in this manuscript are not publicly available because they are subject to a confidentiality agreement with our sponsor (Samsung). The policy of the sponsor restricts open distribution of the data in order to protect proprietary business information. Researchers who wish to obtain access to the data may submit a formal request to the corresponding author. Requests will be evaluated on a case-by-case basis.

Authors' Contributions

AIR, JPFS, and ARPF did the statistical analyses and writing of the article. FAS, AACG, and JAAL performed the writing and review. MRS was responsible for conducting the study, writing, and the review process.

Conflicts of Interest

The authors declare that they have no personal or financial relationships that could be perceived to influence the results or interpretation of this work, except for the research grant provided by Samsung. The authors are not employees, consultants, or shareholders of the sponsor. The work was conducted independently of the commercial interests of the sponsor. The sponsor had no role in study design, data collection, analysis, and manuscript preparation.

References

1. Park J, Seok HS, Kim SS, Shin H. Photoplethysmogram analysis and applications: an integrative review. Front Physiol 2021;12:808451. [doi: 10.3389/fphys.2021.808451] [Medline: 35300400]
2. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
3. Thomson EA, Nuss K, Comstock A, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019 Jun;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
4. Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch™ for measuring maximal heart rate. J Sports Sci 2018 Jul;36(13):1447-1452. [doi: 10.1080/02640414.2017.1397282] [Medline: 29090987]
5. Kim C, Song JH, Kim SH. Validation of wearable digital devices for heart rate measurement during exercise test in patients with coronary artery disease. Ann Rehabil Med 2023 Aug;47(4):261-271. [doi: 10.5535/arm.23019] [Medline: 37536665]
6. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 1988 Jun;124(6):869-871. [doi: 10.1001/archderm.124.6.869] [Medline: 3377516]
7. American College of Sports Medicine. In: Liguori G, Feito Y, Fountaine CJ, Roy B, editors. ACSM's Guidelines for Exercise Testing and Prescription, 11th edition: Wolters Kluwer; 2022.
8. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
9. Jones AM, Doust JH. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. J Sports Sci 1996 Aug;14(4):321-327. [doi: 10.1080/02640419608727717] [Medline: 8887211]
10. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156. [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
11. Costa VAB, Midgley AW, Carroll S, et al. Is a verification phase useful for confirming maximal oxygen uptake in apparently healthy adults? A systematic review and meta-analysis. In: Mourot L, editor. PLoS ONE 2021;16(2):e0247057. [doi: 10.1371/journal.pone.0247057] [Medline: 33596256]
12. Borg GAV. Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise 1982 May;14(5):377. [doi: 10.1249/00005768-198205000-00012]
13. Thompson PD, Arena R, Riebe D, Pescatello LS, American College of Sports Medicine. ACSM's new preparticipation health screening recommendations from ACSM's guidelines for exercise testing and prescription, ninth edition. Curr Sports Med Rep 2013;12(4):215-217. [doi: 10.1249/JSR.0b013e31829a68cf] [Medline: 23851406]
14. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
15. Arifin WN, Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia. A web-based sample size calculator for reliability studies. EIMJ 2018 Sep 28;10(3):67-76. [doi: 10.21315/eimj2018.10.3.8]
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]

XSL•FO
RenderX

17. Hopkins WG. Spreadsheets for analysis of validity and reliability. Sportscience. URL: https://sportsci.org/2015/ValidRely.htm [accessed 2026-03-23]
18. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828. [doi: 10.2196/10828] [Medline: 30855232]
19. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS ONE 2016;11(5):e0154420. [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
20. Lima CS, Bertocco FC, de Oliveira JIV, de Souza TMF, da Silva EP, Von Zuben FJ. Assessment of Samsung Galaxy Watch4 PPG-based heart rate during light-to-vigorous physical activities. IEEE Sens Lett 2024 Jul;8(7):1-4. [doi: 10.1109/LSENS.2024.3408089]
21. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998 Oct;26(4):217-238. [doi: 10.2165/00007256-199826040-00002] [Medline: 9820922]
22. Hajj-Boutros G, Landry-Duval MA, Comtois AS, Gouspillou G, Karelis AD. Wrist-worn devices for the measurement of heart rate and energy expenditure: a validation study for the Apple Watch 6, Polar Vantage V and Fitbit Sense. Eur J Sport Sci 2023 Feb;23(2):165-177. [doi: 10.1080/17461391.2021.2023656] [Medline: 34957939]

Abbreviations

CPET: cardiopulmonary exercise testing
CV: coefficient of variation
ECG: electrocardiogram
GW6: Galaxy Watch 6
HR: heart rate
HRR: heart rate reserve
ICC: intraclass correlation coefficient
LoA: limit of agreement
MAE: median absolute error
MAPE: median absolute percentage error
RER: respiratory exchange ratio
RMSE: root mean square error
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Coristine; submitted 07.Aug.2025; peer-reviewed by D Zubac, GPT Areas; revised version received 10.Feb.2026; accepted 10.Feb.2026; published 16.Apr.2026.

Please cite as:
Inoue A, Soares JPF, Antunes-Santos F, Ferreira A, Gonçalves A, Alcântara JA, dos Santos MR
Heart Rate Estimation Using the Galaxy Watch During Maximal Cardiopulmonary Exercise Testing: Cross-Sectional Validation Study
JMIR Cardio 2026;10:e81917
URL: https://cardio.jmir.org/2026/1/e81917
doi:10.2196/81917

© Allan Inoue, João Paulo Ferreira Soares, Felipe Antunes-Santos, Alexandre Ferreira, Alberto Gonçalves, João Arthur Alcântara, Marcelo Rodrigues dos Santos. Originally published in JMIR Cardio (https://cardio.jmir.org), 16.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cardiac Telerehabilitation Using a Smartwatch and a Gamified Smartphone App: Single-Arm Pre-Post Feasibility Study

Yusaku Arima*, MD; Toshiki Kaihara*, MD, PhD; Megumi Nakamura, MD; Keita Sone, MD; Toshiya Yoshida, MD; Yui Utsugi, MD, PhD; Shiori Takizawa, MD; Shunichi Doi, MD, PhD; Kei Honda, MD; Yoshikuni Kobayashi, MD, PhD; Akira Kasagawa, MD, PhD; Yasuhito Kawagoe, MD; Takahiko Kai, MD, PhD; Masashi Koga, MD; Takumi Higuma, MD, PhD; Yasuhiro Tanabe, MD, PhD; Yoshihiro Akashi, MD, PhD

Department of Cardiology, St. Marianna University School of Medicine, Sugao 2-16-1, Miyamae-ku, Kawasaki, Japan

*these authors contributed equally

Corresponding Author:
Toshiki Kaihara, MD, PhD
Department of Cardiology, St. Marianna University School of Medicine, Sugao 2-16-1, Miyamae-ku, Kawasaki, Japan

Abstract

Background: Home-based cardiac rehabilitation (CR) using digital health technologies (ie, cardiac telerehabilitation [CTR]) has emerged as a practical alternative to conventional center-based CR, particularly during and after the COVID-19 pandemic. However, maintaining sustained participation in CR remains challenging. Gamification holds the potential to enhance motivation and adherence in CR, but its role in CTR for patients with acute coronary syndrome (ACS) remains under-studied.

Objective: This feasibility study evaluated the feasibility, acceptability, and safety of a combination of a gamification-enabled smartphone app and a smartwatch supporting CTR after ACS. We focused specifically on participation motivation, adherence to prescribed exercise intensity, and short-term physiological outcomes.

Methods: This single-arm, pre-post intervention study was conducted at 2 Japanese institutions. Sixteen patients diagnosed with ST-elevated myocardial infarction or non–ST-elevated myocardial infarction and undergoing percutaneous coronary intervention were enrolled after discharge. Each patient received a smartphone and smartwatch connected to the Shin-po Kei app, through which participants earned points when exercising within their target anaerobic threshold heart rate (HR) range (±10 bpm). The 1-month intervention prescribed walking for 30 minutes or longer 3 to 5 days or more per week, at an intensity corresponding to a Borg scale score of 11 to 13. The primary end point was the number of app log-in days divided by the number of intervention days (app use rate). Secondary end points included the target HR exercise adherence rate (defined as the proportion of patients who exercised for 30 minutes within the target HR range on 3 or more days per week), changes in peak oxygen consumption, and changes in perceived exertion using a visual analog scale.

Results: The mean patient age was 62 (SD 13) years, and 94% (15/16) were male. The mean intervention duration was 29 (SD 6) days, and the overall app use rate was 59.9%. The target HR exercise adherence rate was 50%. Peak oxygen consumption increased from 18.4 (SD 4.1) to 21.1 (SD 5.4) mL/kg/min, and the mean visual analog scale score for exercise-related hassle reduced from 57 (SD 33) to 32 (SD 26). No adverse events related to app use were identified, and 1 hospitalization unrelated to the intervention occurred during the study period.

Conclusions: This feasibility study suggests that short-term engagement with a smartwatch-linked gamified smartphone app is achievable in CTR after ACS. However, adherence to predefined gamified exercise standards was lower than app use rates, suggesting that refinement of reward rules and target HR thresholds may be necessary.

(JMIR Cardio 2026;10:e85808) doi:10.2196/85808

KEYWORDS

cardiac telerehabilitation; smartphone app; wearable technology; gamification; digital health; acute coronary syndrome

Introduction

Home-based cardiac rehabilitation (HBCR) is emerging as a feasible and effective alternative to traditional center-based cardiac rehabilitation (CBCR) for patients with cardiovascular disease [1,2]. HBCR achieves improvements in exercise capacity, quality of life, and management of cardiovascular risk factors comparable to those achieved with CBCR while offering enhanced convenience, reduced burden of hospital travel, and improved access, particularly for older adults and residents of rural or medically underserved areas [1]. Furthermore, HBCR is associated with high patient satisfaction and has the potential to reduce health care costs [1,3]. However, the participation rate in CBCR in Japan remains low, with reports indicating that only

XSL•FO
RenderX

approximately 3% to 9% of eligible patients participate, suggesting that barriers, including issues related to accessibility, persist [4,5]. Contributing factors include a shortage of facilities offering cardiac rehabilitation (CR), low awareness among patients and health care providers, and geographical barriers such as remote islands and mountainous regions [6]. These barriers hinder sustained participation in hospital-based programs and limit the reach of conventional CR. To overcome these challenges, HBCR with telemedicine-based approaches—collectively termed cardiac telerehabilitation (CTR)—was developed. CTR uses digital health technologies such as smartphones, wearable devices, videoconferencing, and internet-connected ergometers to provide remote exercise coaching, telemonitoring, patient education, and lifestyle support [7]. Evidence suggests that CTR can achieve outcomes noninferior to traditional CR in terms of compliance, hospitalization rates, and long-term maintenance of physical activity.

The use of smartphone apps (mobile health [mHealth]) in cardiology is rapidly expanding, with promising results. For example, blood pressure management apps have demonstrated significantly superior blood pressure control compared with standard care [8]. Smartwatches recording chest electrocardiograms show high similarity to conventional 12-lead electrocardiograms [9]. A recent randomized controlled trial (RCT) found that a smartphone app linked to a Bluetooth-connected monitoring device improved dyspnea symptoms in patients with heart failure (HF) [10]. However, few studies have evaluated smartphone apps incorporating "gamification" [11]—integration of game elements into CR—for managing coronary artery disease.

A critical challenge is whether users will consistently use the app at the intended frequency. Continuous use is essential for achieving health benefits, but ensuring adherence is difficult. Gamification elements such as points and rewards may enhance motivation and contribute to reducing cardiovascular risk [12]. Smartphone apps incorporating reward systems directly linked to exercise around the anaerobic threshold heart rate (ATHR) have been virtually nonexistent. Although gamification approaches targeting step goals have been reported [13], few studies have incorporated physiological target values obtained from cardiopulmonary exercise testing (CPET) into CR using gamification. Previous research [14] reported a virtual reality–based CR program for patients with HF. This program monitored heart rate (HR) in real time, automatically adjusted exercise intensity, and maintained patients within target HR zones. Another gamification study [15] used avatars representing patients' hearts and tokens that could be earned by addressing risk factors. However, in these studies, the "reward" element—a key component of gamification—was not explicitly linked to achieving specific HR thresholds.

This study's intervention (a smartphone app incorporating gamification elements) is expected to enhance patient behavior change and demonstrate the app's clinical significance by establishing a motivational mechanism. This is achieved by linking the "reward" component of gamification to individualized physiological targets derived from CPET—specifically, the ATHR—thereby integrating it with established principles of exercise prescription. This feasibility study explores the applicability of a gamified smartphone app combining ATHR-based point-earning functionality with HR monitoring via a smartwatch for CR. It specifically focuses on evaluating its acceptability and procedural feasibility as a foundation for future large-scale intervention development.

Methods

Study Design

This prospective single-arm feasibility study aimed to evaluate the usefulness of a gamified smartphone app in CTR. Sixteen patients with ST-elevated myocardial infarction (STEMI) or non-STEMI were enrolled and underwent CTR using a smartphone app and smartwatch immediately after discharge. The app incorporated gamification features, allowing users to earn points for exercise performed within an appropriate HR range and providing feedback based on the points earned.

Recruitment

Patients were recruited from 2 facilities in Kawasaki City, Japan (St. Marianna University Hospital and Kawasaki Municipal Tama Hospital). Eligible patients were those diagnosed with STEMI or non-STEMI based on the criteria of the Japanese Circulation Society 2018 Guideline on Diagnosis and Treatment of Acute Coronary Syndrome [16], who underwent coronary intervention, routinely used a smartphone, and were able to walk outdoors without assistance. All patients received a written explanation of the study protocol and provided written informed consent prior to enrollment. Exclusion criteria included inability to provide consent, severe disease (eg, cancer or collagen vascular disease), or mobility impairment requiring assistance.

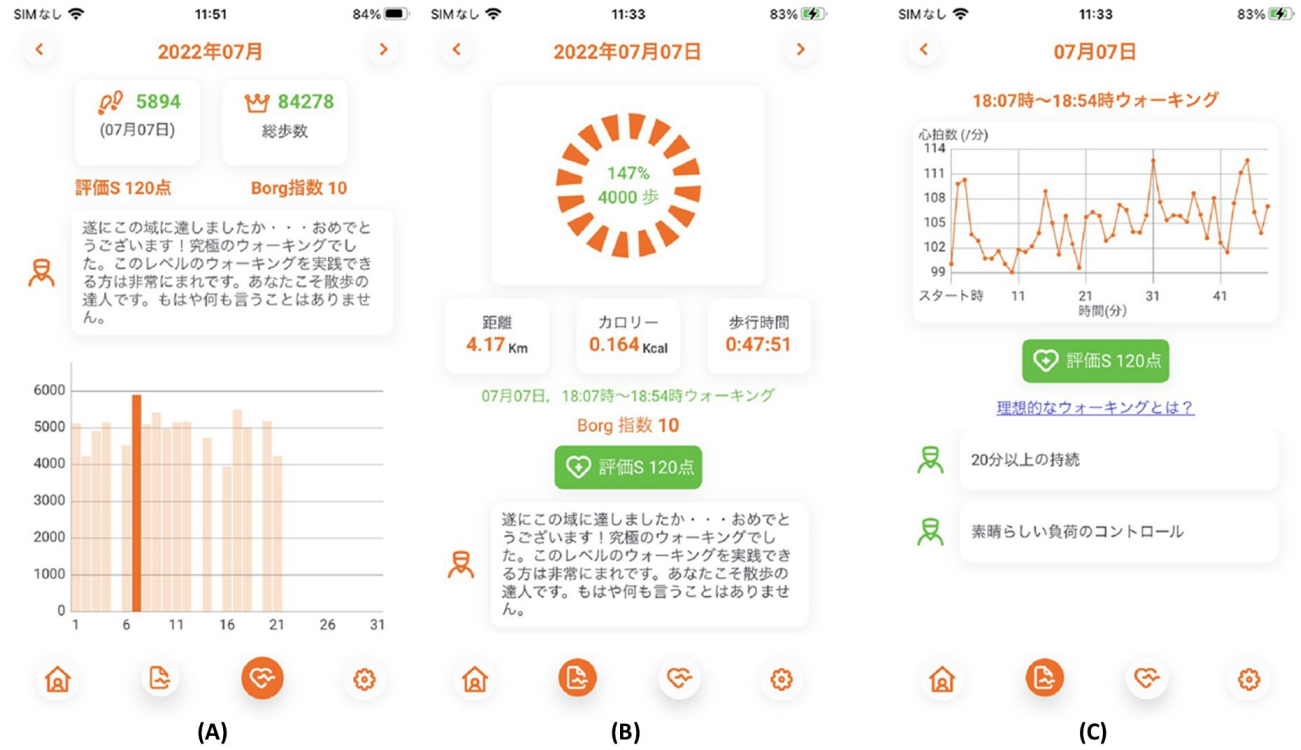
Intervention

Patients in the intervention group participated in the CTR program after discharge. This program included warm-up and cool-down stretches, aerobic exercise, and resistance training, which were explained in advance by physicians and physical therapists prior to discharge. The aerobic exercise component was supported by digital devices. Patients were loaned an iPhone 7 (Apple Inc) and an Apple Watch Series 3 (Apple Inc). The Shin-po Kei app was installed on the smartphone and paired with the smartwatch via Bluetooth. Patients were instructed to perform stretching and resistance training, plus aerobic exercise consisting of walking for at least 30 minutes (Borg scale score 11 - 13) on 3 to 5 days or more per week, based on the Japanese Circulation Society–Japanese Association of Cardiac Rehabilitation 2021 Guideline on Rehabilitation in Patients with Cardiovascular Disease [17]. For the Borg scale, patients manually entered their scores into the app after each exercise session. Since the first day of the study period coincided with the date of discharge and the last day coincided with the first outpatient visit following discharge (1 month later), 2 days were subtracted from the period calculated based on the dates to determine the intervention period. The app recorded HR and exercise duration, awarding points for exercise within the target HR range (ATHR ±10 bpm). Four points were awarded for each minute spent within the target HR range, allowing a maximum of 120 points for a 30-minute exercise session. Points were

XSL•FO
RenderX

deducted at a rate of 2 points per minute for exceeding the target range. These precautions were implemented because patients were immediately after discharge and safety was the top priority. Patients could directly view their daily achievements and progress on the app through automatically generated feedback messages and visual data (eg, line graphs; Figure 1 shows screenshots and Multimedia Appendix 1 provides a translation of the screenshots into English). Regarding step count, although it was displayed in the window, no further feedback was provided. Feedback was provided after completion of the walk (ie, after the patient sent the data via the smartphone app). The underlying logic for the automatic feedback messages was based on points earned (ie, HR during exercise and exercise duration).

Figure 1. Shin-po Kei app screenshots. (A) The first screen shows the number of steps (daily and cumulative), the Borg scale score, and a step count histogram; (B) the second screen displays exercise distance, calories burned, and exercise time; and (C) the third screen shows a line graph with heart rate (bpm) on the y-axis and exercise time (minutes) on the x-axis. Each screen also shows the points earned for the day, along with an automatically generated comment based on the points.



Outcome Measures and Data Collection

The primary end point was the proportion of days during the period between discharge and the first outpatient visit on which an intervention occurred (ie, the patient logged into the app). “App use” was defined as “logging into the app” and was assessed exclusively using app-recorded data automatically collected and verified by the app and smartwatch systems, without reliance on self-reported measures.

Secondary end points included the proportion of patients who exercised for 30 minutes within the target HR range on 3 or more days per week using the smartphone app during the period from discharge to the first outpatient visit (target HR exercise adherence rate), and changes in peak oxygen consumption (VO_2), low-density lipoprotein cholesterol, hemoglobin A_{1c}, and depression or anxiety symptoms (Hospital Anxiety and Depression Scale). Peak VO_2 and anaerobic threshold were measured using symptom-limited CPET and the V-slope method. CPET was performed using a cycle ergometer (SE-8; Mitsubishi Electric Engineering Corp) and a respiratory gas analyzer (Minato Medical Science Co Ltd). The cycle ergometer exercise protocol included a 0-W warm-up and a 10-W/min incremental load. Preintervention CPET was conducted at the start of the study, and postintervention CPET was conducted at the end of the study. Additional secondary end points included changes in exercise-related hassle (visual analog scale [VAS] with 0-100 scoring, where 0 indicated no hassle and 100 indicated extreme hassle), which was designed to evaluate psychological burden and perceived inconvenience associated with exercise rather than physiological symptoms such as pain or shortness of breath; correlations between type D personality (Type D Scale-14) and the primary end points, and the incidence of new complications. This complication assessment was performed by physicians (outpatient physicians) independent of the study investigators. A questionnaire regarding app use was also administered to assess the app’s contribution to exercise, reasons for exercise discontinuation, inconveniences encountered during app use, and potential gamification elements that could enhance motivation for aerobic exercise if incorporated in the future. These usability and feasibility assessments were conducted using app use rates and study-specific questionnaires developed for this pilot study.

Statistical Analysis

Statistical analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing). Data are presented as mean (SD), median (IQR), or percentage. The

XSL•FO
RenderX

chi-square test was used for proportions, paired *t* tests for pre-post comparisons, and Pearson or Spearman correlation coefficients for parametric and nonparametric data, respectively. An intention-to-treat analysis was performed. A 2-tailed $P<.05$ was considered statistically significant. On the basis of prior research [18,19], the pilot study sample size was calculated as 10% of the projected sample size for the larger RCT (150 patients), requiring at least 15 patients. The sample size was determined based on the feasibility of conducting the pilot study, rather than on a power analysis designed to detect a specific effect size.

Ethical Considerations

The protocol, informed consent form, and other research-related documents for this study were approved by the Ethics Committee of St. Marianna University School of Medicine (approval 5512; UMIN 000046777). This study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. Written informed consent was obtained from all patients prior to enrollment. The initial informed consent approved the study procedures without requiring additional consent for secondary analyses. To ensure patient privacy and confidentiality, all study data were anonymized and deidentified prior to analysis. No financial compensation was provided for participation in this study. The manuscript and supplementary materials contain no images or information that could identify individual patients.

Results

From April to August 2022, a total of 36 consecutive patients met the inclusion criteria, of whom 16 patients enrolled. Overall, 8 patients were admitted to St Marianna University Hospital, and 8 patients were admitted to Kawasaki Municipal Tama Hospital. Table 1 shows the baseline characteristics of the patients. The mean age was 62 (SD 13) years, and 94% (n=15) were male. STEMI accounted for 81% (n=13) of cases. The mean intervention period was 29 (SD 6) days, and the mean length of hospital stay was 12 (SD 1) days. Among the 20 patients who declined participation, the most common reason for refusing consent was “not being able to use a smartphone (not having a smartphone).” The results are summarized in Table S1 in Multimedia Appendix 1.

The overall app use rate (number of log-in days or number of intervention days) was 59.9% (282/471 days). The median app use rate was 71.5% (IQR 32.5%-86.5%; detailed data is shown in Table S2 in Multimedia Appendix 1). The rate was 63.1% (149/236 days) in the first half of the intervention period and 56.6% (133/235 days) in the second half (Figure 2) .

XSL•FO
RenderX

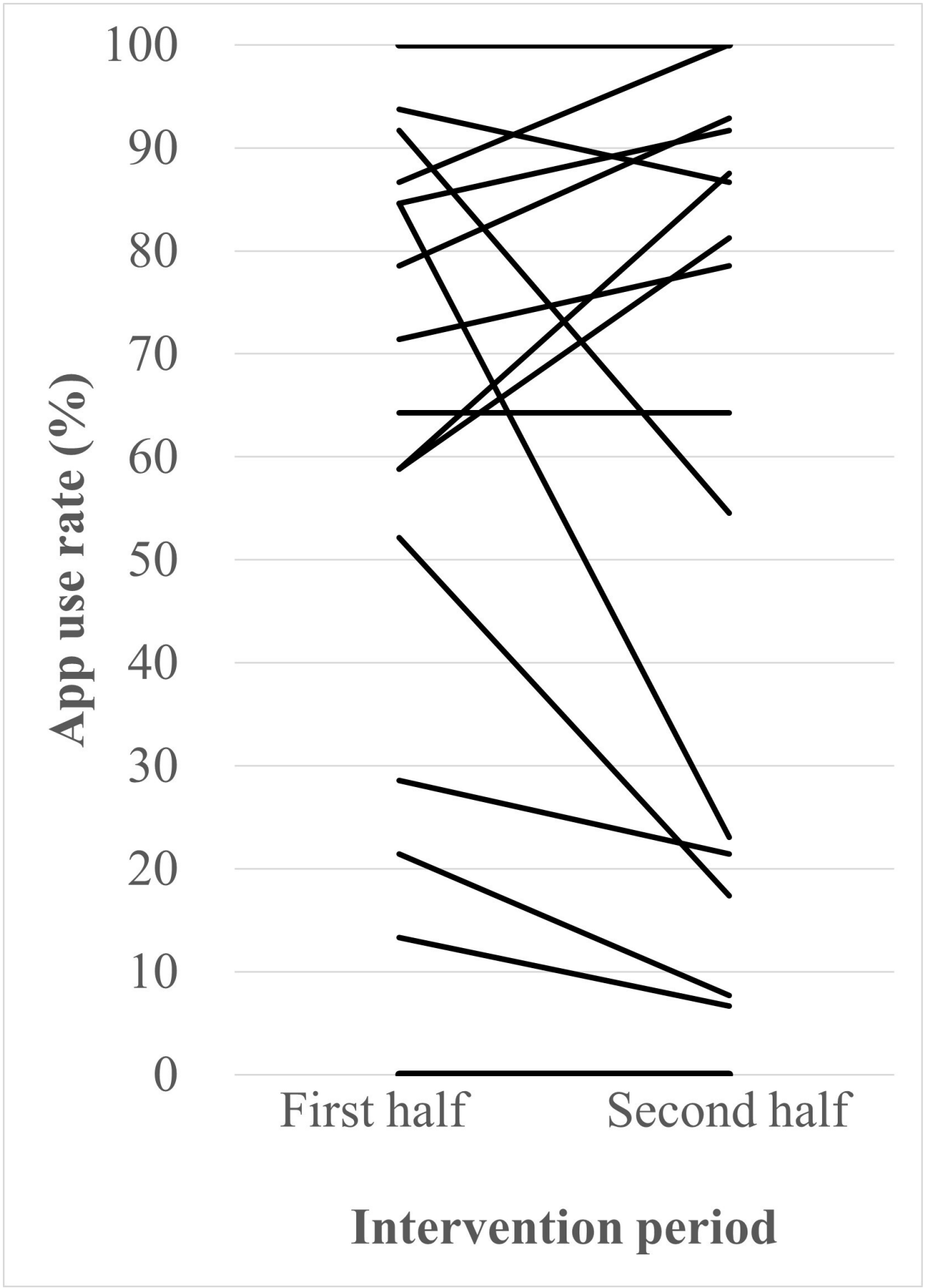
Table . Baseline characteristics of the study participants (N=16).

Characteristic	Participants
Age (years), mean (SD)	62 (13)
Male sex, n (%)	15 (94)
BMI (kg/m^2), mean (SD)	25.4 (4.0)
Smoking status, n (%)	
Current	5 (31)
Former	6 (38)
Never	4 (25)
Not documented[a]	1 (6)
Diagnosis, n (%)	
ST-elevated myocardial infarction	13 (81)
Non–ST-elevated myocardial infarction	3 (19)
Left ventricular ejection fraction ≥50%, n (%)	12 (75)
Laboratory data, median (IQR)	
Low-density lipoprotein cholesterol (mg/dL)	116 (101-127)
Estimated glomerular filtration rate (mL/min/1.73 m^2)	58 (52-69)
Medication data, n (%)	
β-blockers	15 (94)
Angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, or angiotensin receptor neprilysin inhibitors	12 (75)
Statins	16 (100)
Cardiopulmonary exercise testing data, mean (SD)	
Peak oxygen consumption (mL/kg/min)	18.4 (4.1)
Anaerobic threshold heart rate (bpm)	93 (11)
Intervention data, mean (SD)	
Length of hospital stay (days)	12 (1)
Intervention duration (days)	29 (6)

[a]Not recorded in the patient's medical record.

XSL•FO
RenderX

Figure 2. App use rates over time for each patient.



XSL•FO
RenderX

For secondary end points, the target HR exercise adherence rate (defined according to the Japanese Circulation Society–Japanese Association of Cardiac Rehabilitation guidelines [17] and excluding the aforementioned penalties) was 50% (8/16 patients), and the average frequency of exercise sessions lasting 30 minutes was 2.9 (SD 2.1) days per week. The rate of exercise sessions that scored 120 points on the app was 32.9% (155/471 days). Furthermore, the rate of exercise sessions lasting ≥30 minutes within the target HR range, calculated without taking penalties (ie, points deducted for exceeding the target HR range) into account, was 42.0% (198/471 days). Peak VO_2 showed a statistically significant increase from 18.4 (SD 4.1) to 21.1 (SD 5.4) mL/kg/min (P=.004), and the mean VAS score for exercise-related hassle showed a statistically significant reduction from 57 (SD 33) to 32 (SD 26; P=.007). No significant differences were observed in preintervention and postintervention laboratory data (low-density lipoprotein cholesterol and hemoglobin A_{1c}) or Hospital Anxiety and Depression Scale scores. No significant correlation was found between Type D Scale-14 scores and app use rates (r=0.007; P=.98; Figure S2 in Multimedia Appendix 1). One patient did not participate in the intervention, but the data were included in the analysis in accordance with the intention-to-treat principle. One patient was hospitalized during the study period because of severe anemia caused by prolonged prothrombin time-international normalized ratio; however, this event was recognized by the outpatient physician to be the result of inadequate anticoagulant dose adjustment, and no adverse events related to app use were reported (Table S3 in Multimedia Appendix 1).

Survey responses revealed the following: in terms of assessing an app's contribution to exercise, "display of results" (eg, step count, HR, and points) was the most common response (4 patients). In terms of assessing reasons for exercise discontinuation, reasons included "hot weather" (3/16, 19%), "effort required for exercise" (n=3, 19%), and "hospitalization" (n=1, 6%). In terms of assessing inconveniences encountered during app use, 14 patients responded, "none in particular"; 1 patient each cited "discrepancy between points and self-assessment" and "smartwatch battery drain." Regarding potential gamification elements that could enhance motivation for aerobic exercise if incorporated in the future, 14 patients responded. When presented with a list of potential gamification elements, the most frequently selected options were "level system (levels increase with points)," selected by 9 patients, and "calendar display (allows quick daily progress checks in calendar format)," also selected by 9 patients. The next most common responses were "monetary rewards," selected by 6 patients, "nonmonetary rewards," selected by 4 patients, and "character development" (allowing users to improve a character within the app; eg, earn points through exercise therapy to make the character stronger), selected by 4 patients. The detailed responses are shown in Table S4 in Multimedia Appendix 1.

Discussion

Principal Findings

This feasibility study found that the smartphone app use rate (number of log-in days or number of intervention days) was 59.9% over 1 month, suggesting that the gamification app showed moderate short-term uptake during the 4-week intervention; however, causal relationships cannot be inferred from this study. Furthermore, an increase in peak VO_2 was observed during the intervention period, and no significant safety concerns were identified with the use of the gamified app in patients after acute coronary syndrome.

Comparison to Prior Work

Previous studies on digital health, gamification, and CTR [13,15,20] have reported engagement or adherence rates of approximately 30% to 70%. Furthermore, real-world data indicate that the continued use rate of general health apps often drops to approximately 10% after 30 days [21]. Although strict numerical criteria for feasibility were not predefined in this study, the observed 1-month app use rate of 59.9% falls within the range reported in previous studies. However, a tendency toward declining use was observed in the latter part of the intervention period. This aligns with prior findings indicating that sustained engagement with mHealth apps becomes challenging over time. Previous reports attribute this decline to diminishing novelty, competing daily demands, reduced intrinsic motivation, and technical issues such as device synchronization and battery depletion [21]. Our findings may reflect these common barriers to digital health adoption.

To prevent declining app use rates, based on survey findings regarding patient motivation, implementing a calendar view that allows results to be confirmed at a glance and introducing a "level" system—a type of nonmonetary reward obtained by accumulating a certain number of points—may be worth considering. Furthermore, on the basis of survey findings indicating that "visualizing results is important" for supporting the establishment of regular exercise habits, "automatic" result displays may also represent an effective strategy (Multimedia Appendix 1). Studies investigating CR with app or home-based system follow-up have demonstrated improvements in peak VO_2 and health-related quality of life, with some showing significant gains at 1 year and others confirming effects at 12 weeks [22,23]. This study demonstrated similar trends, with improvements in peak VO_2. The observed increase exceeded the minimum clinically meaningful difference of 1 mL/kg/min for reducing mortality risk [24]. Additionally, exercise-related hassle (VAS score) improved.

Gamification

Gamification has been shown to enhance motivation and physical activity levels [13,25], and a recent meta-analysis reported its beneficial effects on glycemic control in type 2 diabetes through improved physical activity [12]. The Shin-po Kei app uses a gamification technique involving earning and losing points based on HR, which may have contributed to the favorable outcomes. This mechanism can be interpreted within a behavioral change framework: points act as nonmonetary

XSL•FO
RenderX

informational rewards that provide immediate feedback and support self-monitoring, feedback, and goal-setting strategies [26]. Meta-analyses show that gamification elements such as points and levels increase physical activity and improve clinical outcomes by enhancing motivation even without financial incentives [27]. These elements deliver continuous feedback that reinforces behavior by fulfilling the core psychological needs of competence and autonomy [28].

CR is fundamentally an intervention designed to promote behavioral change. Although evidence linking reward systems to objective measures of aerobic capacity remains limited, we believe that the introduction of reward systems—a component of gamification also used to promote behavioral change—has high affinity with CR. However, exercise intensity (ie, exercise within target HR zones) was not optimal. This is likely due to the app's safety rules penalizing exercise that exceeded the upper HR limit, which may have been set too low for patients with mild myocardial ischemia (refer to the response regarding inconveniences encountered during app use). This may have contributed to a decrease in patients' exercise adherence rates. Although we stated that "peak VO_2 significantly increased from 18.4 (SD 4.1) mL/kg/min," the actual range was quite broad, from 10.6 to 25.6 mL/kg/min. Gamification rules should be individually set based on clinical stability and physical ability. Rather than eliminating the upper HR limit, future adjustments to the target HR zone should be considered. Expanding the acceptable HR range or introducing dynamically adjusted thresholds based on repeated clinical evaluations may allow greater flexibility while maintaining load management, a fundamental principle of CR. Patient safety remains the paramount priority in any modification to exercise prescriptions based on gamification. Alternatively, applying the system to patient groups for whom exercise exceeding the anaerobic HR threshold poses a low safety risk—such as those with stable angina—could be explored. Such adjustments could maintain safety while enabling more flexible exercise behavior, potentially improving exercise adherence.

Safety

During the 4-week study period, 1 patient was hospitalized, resulting in a temporary interruption of exercise. The cause of hospitalization, as detailed in the Results section, was unrelated to the intervention. No adverse events directly attributable to app use or the exercise prescription were identified, supporting the safety of the mobile CR app. Other gamification studies also report no clear harm [13,15]; therefore, the use of smartphone apps in CR appears to be generally safe. Regarding the relationship between data privacy and ethical issues in mHealth, as explored in this study, concerns such as "secondary data use for commercial purposes" and "psychological constraints arising from monitoring" exist, indicating numerous challenges that require resolution in the future.

Limitations

This feasibility study has several limitations. First, the sample size was small, and the efficacy of gamified smartphone apps should be evaluated in larger RCTs. In the absence of a control group, the observed improvements in peak VO_2 and VAS scores may be attributable, at least in part, to natural physiological recovery during the early postmyocardial infarction phase rather than to the intervention itself.

Second, as mentioned, participation rates were low, potentially introducing selection bias. Furthermore, this study included a predominantly male population (94%), which may limit the generalizability of the findings to female patients. As noted in a recent American Heart Association scientific statement [29], women are known to have lower participation rates in CR programs due to multifactorial barriers. Therefore, caution should be exercised when extrapolating these feasibility findings to female patients. In addition, as "routine smartphone use" was set as the inclusion criterion, patients affected by the digital divide—such as older adults, those without smartphones (the primary reason for exclusion), or those without stable internet access—may have been excluded, potentially limiting the study's generalizability. The same applies to patients with HF; there were almost no patients with concomitant HF in this study, and even when considering the existing study on mHealth [20], the conclusions remain controversial. However, the prior research [30] has highlighted the importance of "digital health readiness" in addressing this issue. This study did not collect data on patients' educational attainment or digital literacy. These factors may influence the feasibility of CTR and the level of engagement with digital health interventions. Future research should consider introducing tools such as digital health readiness questionnaires to further enhance patient engagement.

Third, although the app can be installed on patients' personal iPhones, it was developed exclusively for iOS because of budgetary constraints, which may have affected the generalizability of the findings. To avoid exclusion of patients based on device type, all patients in this study were uniformly provided with an iPhone and an Apple Watch. However, this device-loaning model may be economically challenging to implement in clinical practice. A bring-your-own-device approach, in which patients install the app on their own smartphones, may offer greater cost-effectiveness and user satisfaction. In the subsequent study (the GamiHeart II study), a bring-your-own-device design has been adopted to enhance feasibility and real-world implementability. Furthermore, the devices used in this study were not the latest models at the time of intervention. Therefore, technical inconveniences such as battery drain may partially reflect hardware-related limitations rather than inherent flaws of the software intervention itself. Furthermore, missing data related to hardware functionality may have contributed to the underestimation of exercise adherence based on target HR (Table S2 in Multimedia Appendix 1). Future evaluations using next-generation devices may yield different usability results.

Fourth, although gamification is an integral component of the app used in this study, it is not an isolated feature but is closely intertwined with other functionalities. For example, feedback messages in this study are commonly seen in smartphone apps and vary greatly in nature, ranging from those similar to what physicians provide to patients in routine clinical practice to those designed to enhance motivation for earning monetary or nonmonetary rewards within the app. Separating the gamification elements of this app from other aspects of health care can sometimes be difficult. Although creating an app linked

XSL•FO
RenderX

to more “game-like” elements, such as puzzle games, might allow measurement of gamification capabilities, the results observed in this study likely reflect the combined effect of the entire app, including its gamification features.

Fifth, another important limitation of this study is the relatively short intervention period of approximately 4 weeks. Standard CR programs are typically conducted over 12 weeks or longer; therefore, this study duration is insufficient to adequately evaluate sustained participation or long-term adherence to the intervention. For 1 patient, the hospitalization period was included in the intervention period, so this patient’s adherence period may have been underestimated. Notably, a decline in app use was observed during the latter half of the study period, suggesting that patient engagement may decrease over time. Future, longer-term, large-scale RCTs are warranted to evaluate adherence, cost-effectiveness, and long-term clinical outcomes of gamified CTR.

Finally, usability and acceptability were not evaluated using established, validated measures such as the System Usability Scale or the mHealth App Usability Questionnaire. This limits direct comparisons with digital health interventions in other studies.

Conclusions

This feasibility study suggests that short-term engagement with mHealth using a gamified smartphone app for CTR after acute coronary syndrome is achievable. However, adherence to predefined gamified exercise criteria was lower than app use rates, indicating a potential need for refinement of reward rules and target HR thresholds. These findings warrant further evaluation in larger-scale studies.

Acknowledgments

The authors thank the patients who participated in this study for their dedication and commitment. The authors also thank the medical and nursing staff at St. Marianna University Hospital and Kawasaki Municipal Tama Hospital for their support in patient recruitment and care. The authors are also grateful to the developers of the Shin-po Kei app for their collaboration in adapting the software for this study. Finally, the authors thank the study sponsor (St. Marianna University Hospital) for financial support and assistance in ensuring the successful completion of this research. During the writing of this paper, ChatGPT (GPT-4o; OpenAI) was used for supplementary support in English-language expression. The generated text was ultimately reviewed for accuracy and appropriateness by the authors (Y Arima and T Kaihara), who take full responsibility for the content of this manuscript.

Funding

This work was funded by St Marianna University Hospital.

Data Availability

All datasets analyzed during this study are included in this manuscript.

Authors' Contributions

T Kaihara contributed to the conception or design of the work. Y Arima, MN, KS, TY, YU, ST, SD, KH, Y Kobayashi, AK, Y Kawagoe, T Kai, and MK contributed to data acquisition. T Kaihara contributed to data analysis and interpretation. Y Arima drafted the manuscript. All authors (Y Arima, T Kaihara, MN, KS, TY, YU, ST, SD, KH, Y Kobayashi, AK, Y Kawagoe, T Kai, MK, TH, YT, and Y Akashi) gave final approval of the manuscript and agree to be accountable for all aspects of the work, ensuring its integrity and accuracy. The first (Y Arima) and second (T Kaihara) authors contributed equally to the writing process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables detailing consent refusal reasons, patient app use rates, adverse events, questionnaire responses on exercise motivation, and a translation and summary of the screenshots in Figure 1.
[DOCX File, 29 KB - cardio_v10i1e85808_app1.docx]

References

1. McDonagh ST, Dalal H, Moore S, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2023 Oct 27;10(10):CD007130. [doi: 10.1002/14651858.CD007130.pub5] [Medline: 37888805]
2. Snoek JA, Prescott EI, van der Velde AE, et al. Effectiveness of home-based mobile guided cardiac rehabilitation as alternative strategy for nonparticipation in clinic-based cardiac rehabilitation among elderly patients in Europe: a randomized clinical trial. JAMA Cardiol 2021 Apr 1;6(4):463-468. [doi: 10.1001/jamacardio.2020.5218] [Medline: 33112363]
3. Scherrenberg M, Zeymer U, Schneider S, et al. EU-CaRE study: could exercise-based cardiac telerehabilitation also be cost-effective in elderly? Int J Cardiol 2021 Oct 1;340:1-6. [doi: 10.1016/j.ijcard.2021.08.024] [Medline: 34419529]

XSL•FO
RenderX

4. Kamiya K, Yamamoto T, Tsuchihashi-Makaya M, et al. Nationwide survey of multidisciplinary care and cardiac rehabilitation for patients with heart failure in Japan - an analysis of the AMED-CHF study. Circ J 2019 Jun 25;83(7):1546-1552. [doi: 10.1253/circj.CJ-19-0241] [Medline: 31189753]
5. Kanaoka K, Soeda T, Terasaki S, et al. Current status and effect of outpatient cardiac rehabilitation after percutaneous coronary intervention in Japan. Circ Rep 2021 Feb 27;3(3):122-130. [doi: 10.1253/circrep.CR-20-0143] [Medline: 33738344]
6. Kaihara T, Scherrenberg M, Falter M, et al. Cardiac telerehabilitation - a solution for cardiovascular care in Japan. Circ Rep 2021 Oct 29;3(12):733-736. [doi: 10.1253/circrep.CR-21-0126] [Medline: 34950799]
7. Scherrenberg M, Falter M, Abreu A, et al. Standards for cardiac telerehabilitation. Eur Heart J 2025 Oct 7;46(38):3714-3737. [doi: 10.1093/eurheartj/ehaf408] [Medline: 40742158]
8. Kario K, Nomura A, Harada N, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021 Oct 21;42(40):4111-4122. [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
9. van der Zande J, Strik M, Dubois R, et al. Using a smartwatch to record precordial electrocardiograms: a validation study. Sensors (Basel) 2023 Feb 25;23(5):2555. [doi: 10.3390/s23052555] [Medline: 36904759]
10. Yoon M, Lee S, Choi JY, et al. Effectiveness of a smartphone app-based intervention with Bluetooth-connected monitoring devices and a feedback system in heart failure (SMART-HF trial): randomized controlled trial. J Med Internet Res 2024 Apr 29;26:e52075. [doi: 10.2196/52075] [Medline: 38683665]
11. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining “gamification”. In: MindTrek ’11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments: Association for Computing Machinery; 2011:9-15.
12. Kaihara T, Intan-Goey V, Scherrenberg M, et al. Impact of gamification on glycaemic control among patients with type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. Eur Heart J Open 2021;1(3):oeab030. [doi: 10.1093/ehjopen/oeab030] [Medline: 35919887]
13. Xu L, Tong Q, Zhang X, et al. Smartphone-based gamification intervention to increase physical activity participation among patients with coronary heart disease: a randomized controlled trial. J Telemed Telecare 2024 Oct;30(9):1425-1436. [doi: 10.1177/1357633X221150943] [Medline: 36794484]
14. Blasco-Peris C, Alcolea Garrido JP, Seguí B, et al. Exergaming system for exercise-based cardiac rehabilitation in patients with heart failure: development and usability assessment study of a device prototype. JMIR Serious Games 2025 Jul 16;13:e71385. [doi: 10.2196/71385] [Medline: 40669044]
15. Gallagher R, Chow CK, Parker H, et al. The effect of a game-based mobile app “MyHeartMate” to promote lifestyle change in coronary disease patients: a randomized controlled trial. Eur Heart J Digit Health 2022;4(1):33-42. [doi: 10.1093/ehjdh/ztac069] [Medline: 36743873]
16. Kimura K, Kimura T, Ishihara M, et al. JCS 2018 guideline on diagnosis and treatment of acute coronary syndrome. Circ J 2019 Apr 25;83(5):1085-1196. [doi: 10.1253/circj.CJ-19-0133] [Medline: 30930428]
17. Makita S, Yasu T, Akashi YJ, et al. CORRIGENDUM: JCS/JACR 2021 guideline on rehabilitation in patients with cardiovascular disease. Circ J 2023;87(6):937. [doi: 10.1253/circj.CJ-66-0214] [Medline: 37225499]
18. Connelly LM. Pilot studies. Medsurg Nurs 2008 Dec;17(6):411-412. [Medline: 19248407]
19. Lackey NR, Wingate AL. The pilot study: one key to research success. In: Brink PJ, Wood MJ, editors. Advanced Design in Nursing Research, 2nd edition: SAGE Publications; 1998:375-386.
20. Yokota T, Fukushima A, Tsuchihashi-Makaya M, et al. The AppCare-HF randomized clinical trial: a feasibility study of a novel self-care support mobile app for individuals with chronic heart failure. Eur Heart J Digit Health 2023;4(4):325-336. [doi: 10.1093/ehjdh/ztad032] [Medline: 37538146]
21. Shrestha P. Checking the pulse of health & fitness tracker apps in 2023. Adjust. 2023. URL: https://www.adjust.com/blog/health-tracker-installs-and-retention-data/ [accessed 2026-03-19]
22. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792. [doi: 10.1177/2047487320905717] [Medline: 32106713]
23. Batalik L, Dosbaba F, Hartman M, Batalikova K, Spinar J. Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: a randomized controlled trial. Medicine (Baltimore) 2020 Mar;99(11):e19556. [doi: 10.1097/MD.0000000000019556] [Medline: 32176113]
24. Keteyian SJ, Brawner CA, Ehrman JK, et al. Reproducibility of peak oxygen uptake and other cardiopulmonary exercise parameters: implications for clinical trials and clinical practice. Chest 2010 Oct;138(4):950-955. [doi: 10.1378/chest.09-2624] [Medline: 20522572]
25. Fanaroff AC, Patel MS, Chokshi N, et al. Effect of gamification, financial incentives, or both to increase physical activity among patients at high risk of cardiovascular events: the BE ACTIVE randomized controlled trial. Circulation 2024 May 21;149(21):1639-1649. [doi: 10.1161/CIRCULATIONAHA.124.069531] [Medline: 38583084]
26. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 4;24(1):e26779. [doi: 10.2196/26779] [Medline: 34982715]

XSL•FO
RenderX

27. Edwards EA, Lumsden J, Rivas C, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447. [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
28. Berglund A, Jaarsma T, Berglund E, Strömberg A, Klompstra L. Understanding and assessing gamification in digital healthcare interventions for patients with cardiovascular disease. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):630-638. [doi: 10.1093/eurjcn/zvac048] [Medline: 35709297]
29. Coutinho T, Khadanga S, Adedinsewo D, et al. Cardiac rehabilitation in women: a scientific statement from the American Heart Association. Circulation 2025 Nov 11;152(19):e376-e390. [doi: 10.1161/CIR.0000000000001379] [Medline: 41084821]
30. Scherrenberg M, Falter M, Kaihara T, et al. Development and internal validation of the Digital Health Readiness Questionnaire: prospective single-center survey study. J Med Internet Res 2023 Mar 10;25:e41615. [doi: 10.2196/41615] [Medline: 36897627]

Abbreviations

ATHR: anaerobic threshold heart rate
CBCR: center-based cardiac rehabilitation
CPET: cardiopulmonary exercise testing
CR: cardiac rehabilitation
CTR: cardiac telerehabilitation
HBCR: home-based cardiac rehabilitation
HF: heart failure
HR: heart rate
mHealth: mobile health
RCT: randomized controlled trial
STEMI: ST-elevated myocardial infarction
VAS: visual analog scale
VO_2: oxygen consumption

Edited by A Coristine; submitted 13.Oct.2025; peer-reviewed by H Mautong, J Simmich; revised version received 06.May.2026; accepted 07.May.2026; published 11.Jun.2026.

Please cite as:
Arima Y, Kaihara T, Nakamura M, Sone K, Yoshida T, Utsugi Y, Takizawa S, Doi S, Honda K, Kobayashi Y, Kasagawa A, Kawagoe Y, Kai T, Koga M, Higuma T, Tanabe Y, Akashi Y
Cardiac Telerehabilitation Using a Smartwatch and a Gamified Smartphone App: Single-Arm Pre-Post Feasibility Study
JMIR Cardio 2026;10:e85808
URL: https://cardio.jmir.org/2026/1/e85808
doi:10.2196/85808

© Yusaku Arima, Toshiki Kaihara, Megumi Nakamura, Keita Sone, Toshiya Yoshida, Yui Utsugi, Shiori Takizawa, Shunichi Doi, Kei Honda, Yoshikuni Kobayashi, Akira Kasagawa, Yasuhito Kawagoe, Takahiko Kai, Masashi Koga, Takumi Higuma, Yasuhiro Tanabe, Yoshihiro Akashi. Originally published in JMIR Cardio (https://cardio.jmir.org), 11.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of the Cardio-Meds Mobile App on Heart Failure Knowledge and Medication Adherence: Pilot Randomized Controlled Trial

Victor Buswell[1], BSc; Emmanuelle Massie[2], MD; Elena Tessitore[2], MD; Lisa Simioni[1], MSc; Guillaume Guebey[1], BSc; Hamdi Hagberg[3], BSc; Aurélie Schneider-Paccot[2], BSc; Samaksha Pant[4], MD; Katherine Blondon[5], MD, PhD; Liliane Gschwind[6], PhD; Frederic Ehrler[3], PhD; Philippe Meyer[2], MD

[1]Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2]Heart Failure and Cardiovascular Rehabilitation Unit, Cardiology Service, Geneva University Hospitals, Rue Gabrielle Perret-Gentil 4, Geneva, Switzerland

[3]Computer Science Department, University Hospital of Geneva, Geneva, Switzerland

[4]Cardiology Division, Hôpital de la Tour, Meyrin, Switzerland

[5]Medical Directorate, University Hospital of Geneva, Geneva, Switzerland

[6]Pharmacy Department, University Hospital of Geneva, Geneva, Switzerland

Corresponding Author:
Philippe Meyer, MD
Heart Failure and Cardiovascular Rehabilitation Unit, Cardiology Service, Geneva University Hospitals, Rue Gabrielle Perret-Gentil 4, Geneva, Switzerland

Abstract

Background: Heart failure (HF) is a prevalent chronic condition for which optimal management depends not only on guideline-directed medical therapy but also on patients' understanding of their disease, recognition of warning signs, and sustained medication adherence, which remains challenging in routine care. Mobile health interventions may support therapeutic education and self-management; however, many available apps lack validated content and local relevance. Cardio-Meds is a mobile app developed at Geneva University Hospitals to support HF self-management through structured educational content, interactive quizzes, medication lists with reminders, and tools for monitoring weight and vital signs.

Objective: This study aims to evaluate the impact of a 30-day Cardio-Meds intervention on HF knowledge and medication adherence in patients with HF with reduced or mildly reduced ejection fraction.

Methods: We conducted a single-center, pilot randomized controlled trial in patients followed at the outpatient HF clinic or enrolled in cardiac rehabilitation at Geneva University Hospitals in 2024. Eligible participants had HF with a left ventricular ejection fraction less than 50%, were receiving HF-specific pharmacotherapy, speak French, and owned a smartphone. Participants were recruited by phone and randomized to Cardio-Meds use for 30 days, a self-guided intervention with a single standardized technical support call. Outcomes were self-assessed using standardized questionnaires: HF knowledge and self-management using the Dutch Heart Failure Knowledge Scale (DHFKS; score range 0 - 15); medication adherence using the Basel Assessment of Adherence to Immunosuppressive Medication Scale, covering initiation, implementation, and persistence; and usability in the intervention group using the System Usability Scale (score range 0 - 100). Between-group differences in DHFKS scores were analyzed using analysis of covariance adjusted for baseline values.

Results: A total of 49 participants were included (25 intervention, 24 control); 78% (n=38) were male, and the mean age was 62 (SD 11.4) years. In the intervention group, median app usage was 123 (IQR 74 - 273) minutes, with a median of 43 (IQR 19 - 85) logins. Mean baseline DHFKS scores were similar between groups (intervention 11.1, SD 2.4 vs control 10.5, SD 2.9). At 30 days, mean scores increased significantly in the intervention group (12.4, SD 2.4; mean change +1.3; P<.001) and remained stable in the control group (10.4, SD 3; mean change –0.1; P=.82), with a significant adjusted between-group difference of +1.3 points (P<.001). No significant between-group differences were observed for medication adherence. Usability was high, with a mean score of 84.3 (SD 15), and 64% (16/25) of intervention participants reported that they would continue using the app.

Conclusions: In a stable ambulatory HF population, the Cardio-Meds intervention demonstrated short-term improvement in HF knowledge, while no effect was observed on medication adherence within the 30-day follow-up period. The app showed high usability and acceptability. Larger multicenter studies with longer follow-up are needed to assess clinical impact.

(JMIR Cardio 2026;10:e83022) doi:10.2196/83022

XSL•FO
RenderX

KEYWORDS

heart failure; mHealth; medication adherence; patient education; digital health intervention; randomized controlled trial; mobile health

Introduction

Heart failure (HF) is a chronic disease affecting more than 64 million people worldwide [1], including approximately 200,000 in Switzerland [2]. Its prevalence is increasing due to aging populations and improved survival. Despite advances in therapy, HF remains a leading cause of morbidity, mortality, and health care utilization. In the European Society of Cardiology-HF Long-Term Registry, 1-year all-cause mortality after hospitalization for acute HF was 23.6%, with a combined incidence of death or HF-related readmission reaching 36% [3]. Readmissions contribute significantly to the clinical and economic burden of HF. In Europe and North America, HF accounts for 1% to 2% of total health care expenditures, largely due to hospitalizations [4,5]. Effective outpatient strategies are therefore crucial. Guideline-directed therapy for patients with reduced or mildly reduced ejection fraction (HFrEF and HFmrEF) includes 4 foundational drug classes: angiotensin receptor–neprilysin inhibitors, angiotensin receptor blockers or angiotensin-converting enzyme inhibitors, beta-blockers, mineralocorticoid receptor antagonists, and sodium-glucose cotransporter 2 inhibitors [6]. These treatments clearly improve outcomes but require a high level of adherence, which remains challenging in patients with multiple comorbidities.

Therapeutic education and structured follow-up are therefore central to improving adherence and clinical outcomes in patients with HF. Cardiac rehabilitation programs, recommended by both European and American guidelines (class I, level A), reduce readmissions and improve quality of life [7]. Broader evidence also supports educational interventions in improving self-care behaviors and adherence [8-10]. A meta-analysis by Van Spall et al [8] demonstrated reduced readmissions with nurse-led follow-up and therapeutic education, while Ruppar et al [10] found lower mortality with adherence-focused interventions. However, scalability and patient engagement remain challenges. Mobile health (mHealth) technologies offer accessible, personalized, and cost-effective support. Self-management apps provide education, monitoring, and behavioral reinforcement outside clinical settings. The 2021 European Society of Cardiology guidelines recommend self-management strategies to reduce HF-related hospitalizations and mortality [11,12].

Recent randomized controlled trials (RCTs) have evaluated multicomponent mHealth interventions integrating mobile apps and connected devices, reporting improvements in self-care behaviors and symptom outcomes in HF populations. Kitsiou et al [13] tested the iCardia4HF program in a phase 1 RCT of combined consumer apps and feedback messages, demonstrating feasibility and preliminary efficacy on self-care measures. Moreover, the SMART-HF study employed a smartphone app with remote monitoring and feedback, showing symptomatic benefits and supporting the role of digital strategies in HF management [14]. Systematic reviews of recent RCTs also highlight the potential of digital health interventions to affect clinical outcomes and self-care, though evidence remains heterogeneous [15].

However, current HF apps often lack validated content and integration into national health care systems. Dunn Lopez et al [16] found that most apps were difficult to understand, poorly referenced, and lacked personalized feedback. Apps such as WOW ME 2000mg, WebMD, and My Cardiac Coach provide generic tools but do not account for local medications or structures [17-19]. Usability, health literacy, and personalization remain key success factors for mobile apps [20,21].

This RCT evaluates the impact of Cardio-Meds, an mHealth app developed by the Cardiology Department at Geneva University Hospitals (HUG) [22], on HF knowledge and therapeutic adherence in patients with HFrEF or HFmrEF. The study aims to determine whether a mobile intervention can enhance disease understanding and support adherence in a real-world outpatient setting.

Methods

Study Design

This single-center, prospective, pilot RCT was conducted at HUG. Participants were randomly assigned to either the intervention group (using the Cardio-Meds app) or the control group (usual care). A total of 50 participants were planned for inclusion, with 25 participants per group. The primary objective was to assess whether Cardio-Meds improves knowledge and self-management of HF compared with usual care. Secondary and exploratory objectives were to evaluate its impact on medication adherence and user satisfaction. The study was designed as an exploratory evaluation of short-term effects on HF knowledge, feasibility, and usability rather than as a definitive efficacy trial for medication adherence.

Ethical Considerations

The study was submitted to the local ethics committee (Commission cantonale d’éthique de la recherche, Geneva, Switzerland; ID 2023 - 01337) for review. The Commission cantonale d’éthique de la recherche reviewed the submission (ID 2023 - 01337) and determined that formal approval was not required given the low-risk, educational nature of the study and the absence of clinical or safety end points. All participants provided written informed consent prior to participation. Participant privacy and confidentiality were maintained. Study data were securely stored at Geneva University Hospitals and analyzed in pseudonymized form. Participants did not receive any financial compensation for participation.

This exploratory, single-center RCT was not prospectively registered in a public trial registry. At the time of study initiation, prospective registration was not requested for this type of low-risk educational intervention. All study outcomes, analyses, and end points were predefined before data collection and are reported consistently, with no outcome switching. Future

XSL•FO
RenderX

larger-scale or multicenter randomized trials evaluating Cardio-Meds will be prospectively registered in accordance with international recommendations.

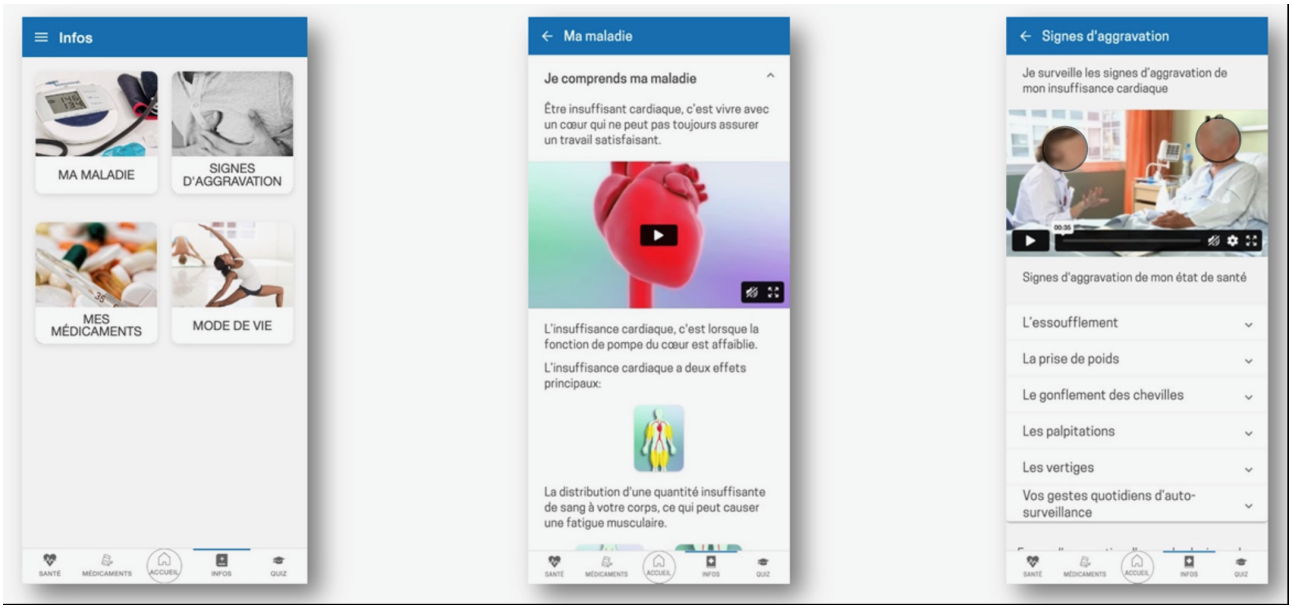
Description of the Intervention (Cardio-Meds App)

Cardio-Meds is a research version of a mHealth app developed at HUG through collaboration between cardiology, pharmacy, the medical directorate, and IT departments. A recent usability study rated it good to excellent, supporting its feasibility in the HF population [23]. The app was designed with and for patients to help individuals with HF manage their condition through educational resources, self-monitoring, and adherence support.

Key features include the following:

- Educational content and daily quizzes on HF, its treatment, and lifestyle. The content is organized into 4 chapters (my condition, warning signs, my medication, and lifestyle, Figure 1). Quizzes provide immediate feedback and brief explanations (Figure 2).
- Tools to track weight, blood pressure, and heart rate, enabling early detection of decompensation. Graphs can be generated and used during consultations (Figure 3).
- A treatment plan, easily accessible on the user's phone, allows patients to enter their medication regimen (Figure 4). Medications can be added by scanning the barcode, enabling the app to identify the drug via access to the Swiss medication database. The user can then specify the dosage and timing for each medication intake (Figure 5).
- Intake confirmation and optional reminders to support adherence (Figures 3, 4, and 6).

Figure 1. Educational information section of the Cardio-Meds mobile app for heart failure self-management. Screenshots show the structured educational content available in the Cardio-Meds app, organized into thematic modules (disease understanding, warning signs, medications, and lifestyle). The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024.

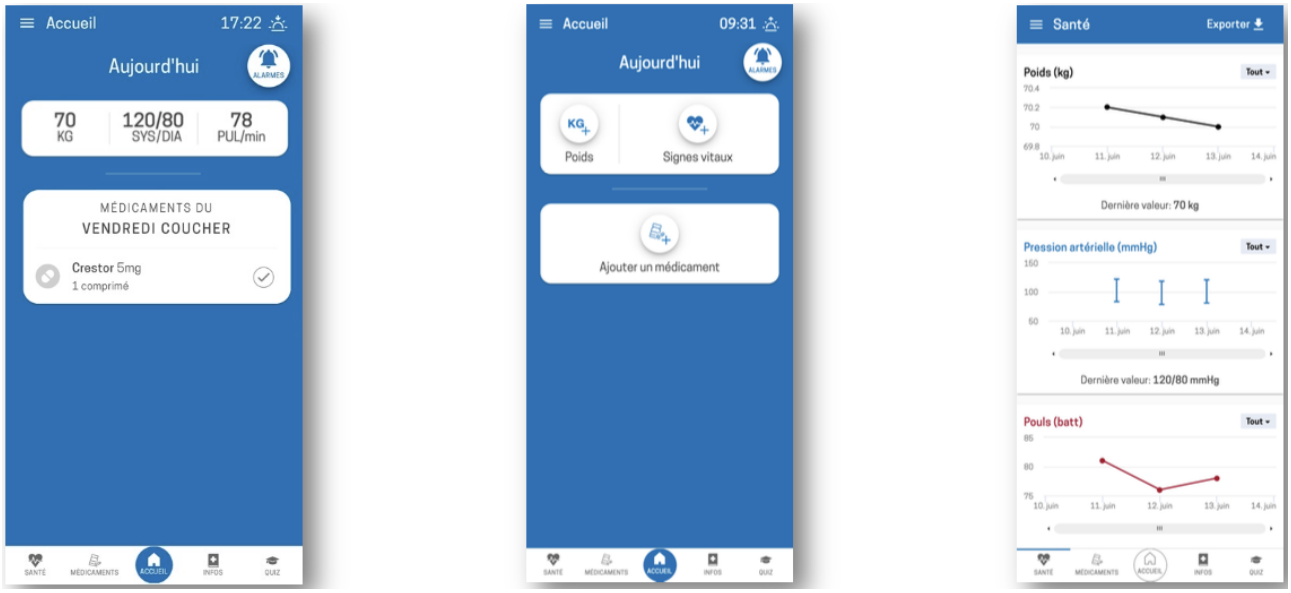


XSL•FO
RenderX

Figure 2. Interactive quiz module of the Cardio-Meds mobile app for heart failure self-management. Screenshots depict the quiz interface of Cardio-Meds, including multiple-choice questions and immediate feedback with brief educational explanations. The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024.



Figure 3. Home screen and health monitoring section of the Cardio-Meds mobile app**.** Screenshots illustrate the home interface and the health monitoring module of Cardio-Meds app. Shown features include daily medication overview, intake confirmation, and graphical tracking of self-reported weight, blood pressure, and heart rate. The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024. All data displayed in the app screenshots are fictitious and were entered solely for illustrative purposes.



XSL•FO
RenderX

Figure 4. Medication management section of the Cardio-Meds mobile app (treatment list, intake confirmation, and history). Screenshots illustrate the medication section of Cardio-Meds, including the individualized treatment list, daily intake confirmation, and medication history. The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024. All data displayed in the app screenshots are fictitious and were entered solely for illustrative purposes.

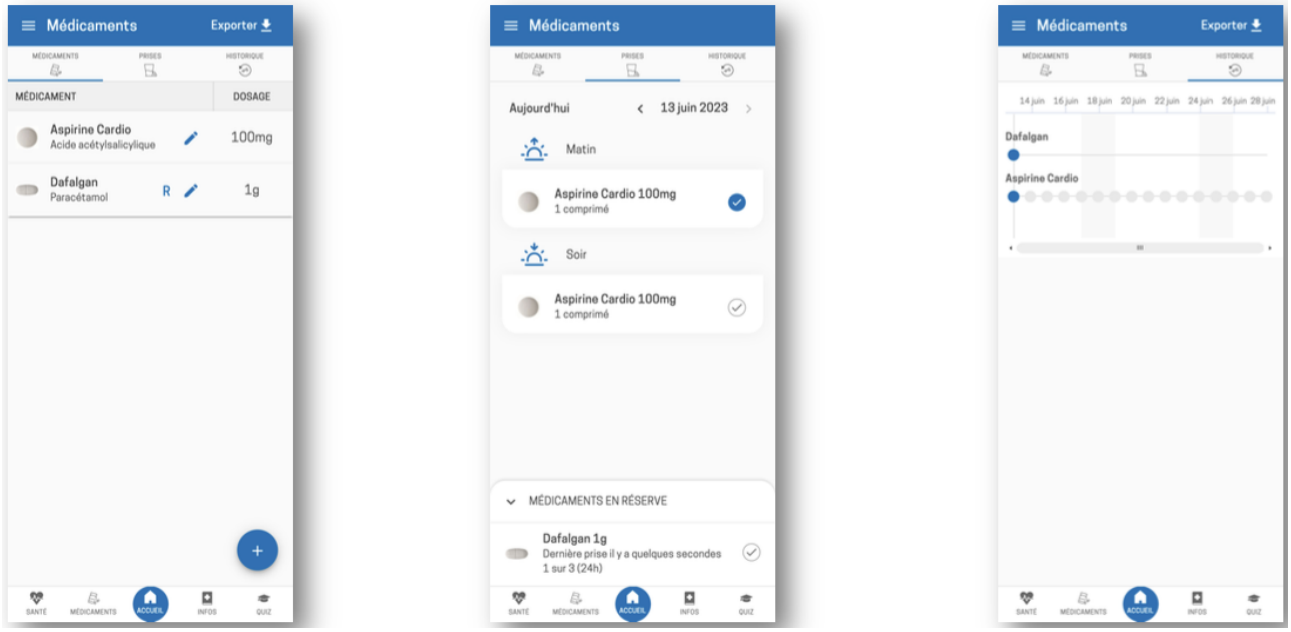
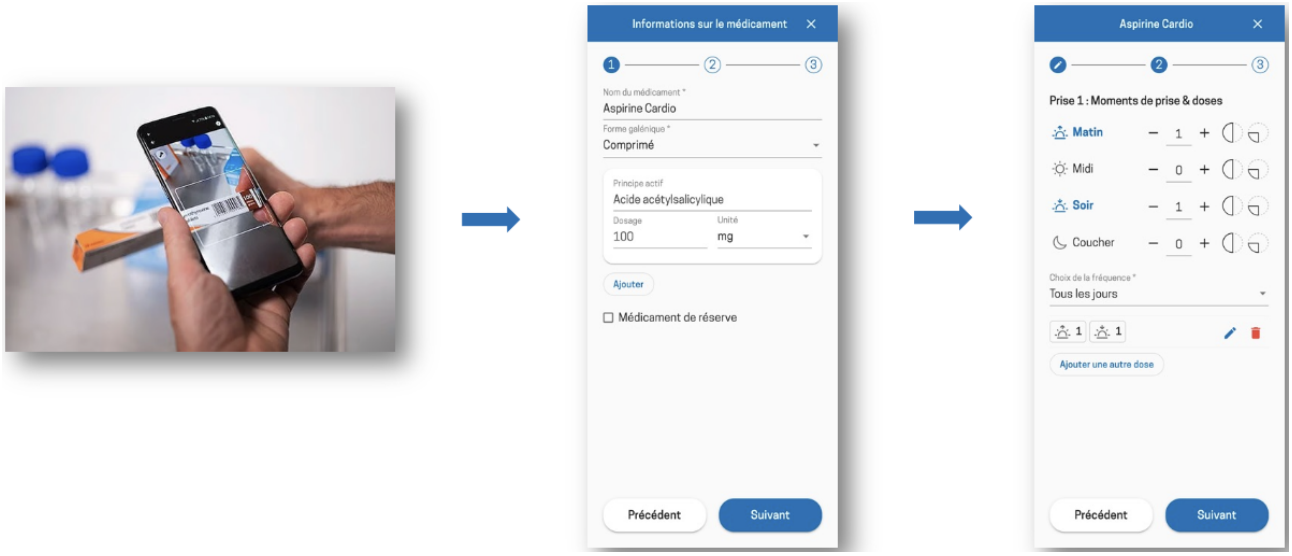
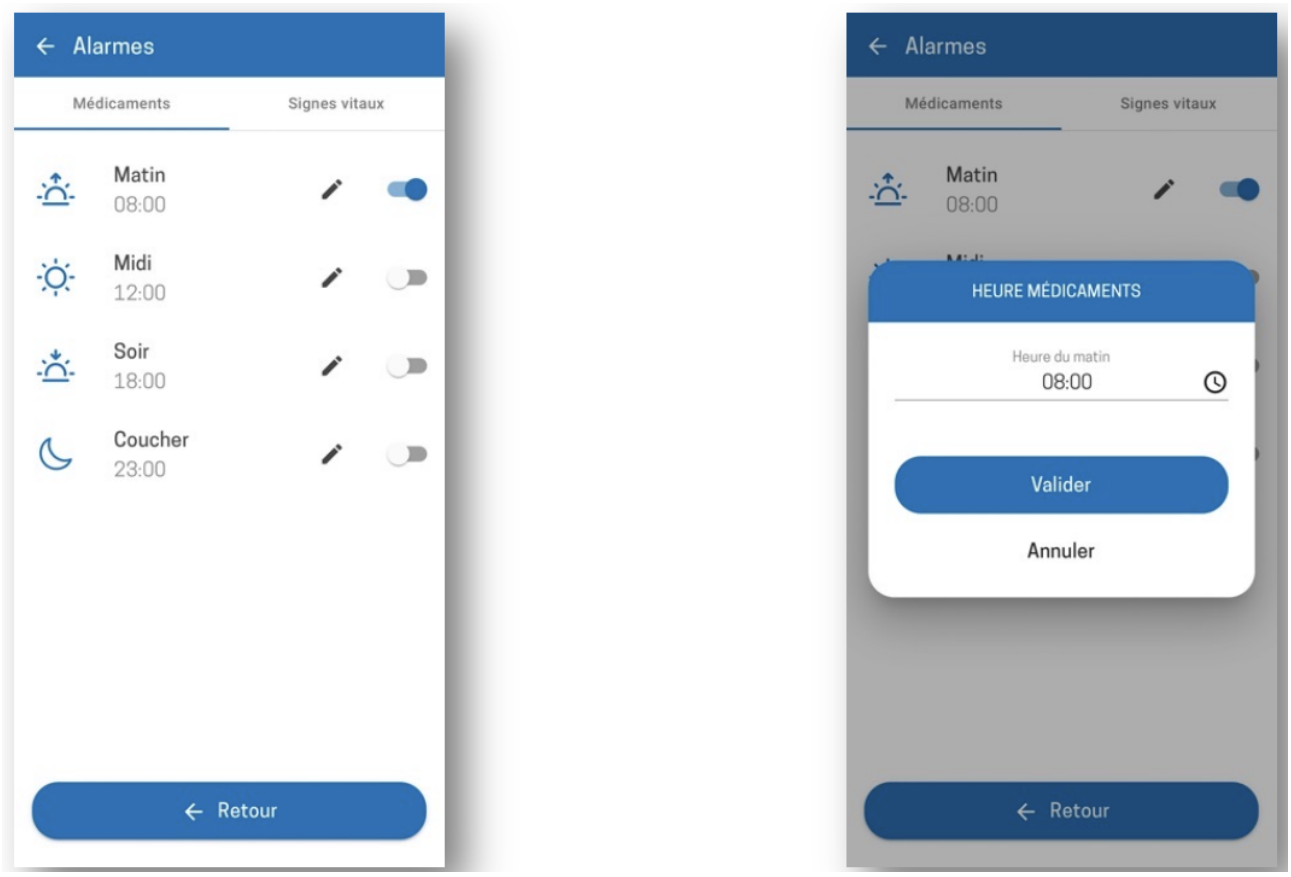


Figure 5. Procedure for entering a medication into the Cardio-Meds mobile app. Screenshots show the step-by-step process for adding a medication in Cardio-Meds, including barcode scanning linked to the Swiss medication database, dosage specification, and intake scheduling. The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024.



XSL•FO
RenderX

Figure 6. Procedure for setting medication reminder alarms in the Cardio-Meds mobile app. Screenshots illustrate the configuration of medication reminder alarms, including timing and activation of notifications. These reminders are optional. The app was evaluated in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult outpatients with heart failure during a 30-day intervention period in 2024.



The Cardio-Meds app enables patients to manage health data, access structured learning, and share information with health care providers. It was designed to ensure ease of use and fulfill legal and best practice requirements for data protection and confidentiality.

Study Population and Recruitment

Participants were recruited from the HF clinic and cardiac rehabilitation programs at HUG from March to November 2024. Two cardiologists and 2 specialized HF nurses preidentified eligible participants, who were contacted by a medical student who explained the study and verified smartphone ownership. Participation was voluntary, with the option to withdraw at any time without affecting medical care.

Participants were eligible for inclusion if they met the following criteria:

- Aged 18 years or older
- Diagnosed with HF with reduced (≤40%, or HFrEF) or mildly reduced (41% - 49% or HFmrEF) left ventricular ejection fraction
- Currently receiving HF-specific medications
- No cognitive impairment and fully capable of decision-making
- Capability to read, understand, and communicate in French
- Ownership of a smartphone

Participants were excluded if they met any of the following conditions:

- Inability or unwillingness to comply with study procedures
- Status as an asylum seeker, individuals experiencing homelessness, or detainees

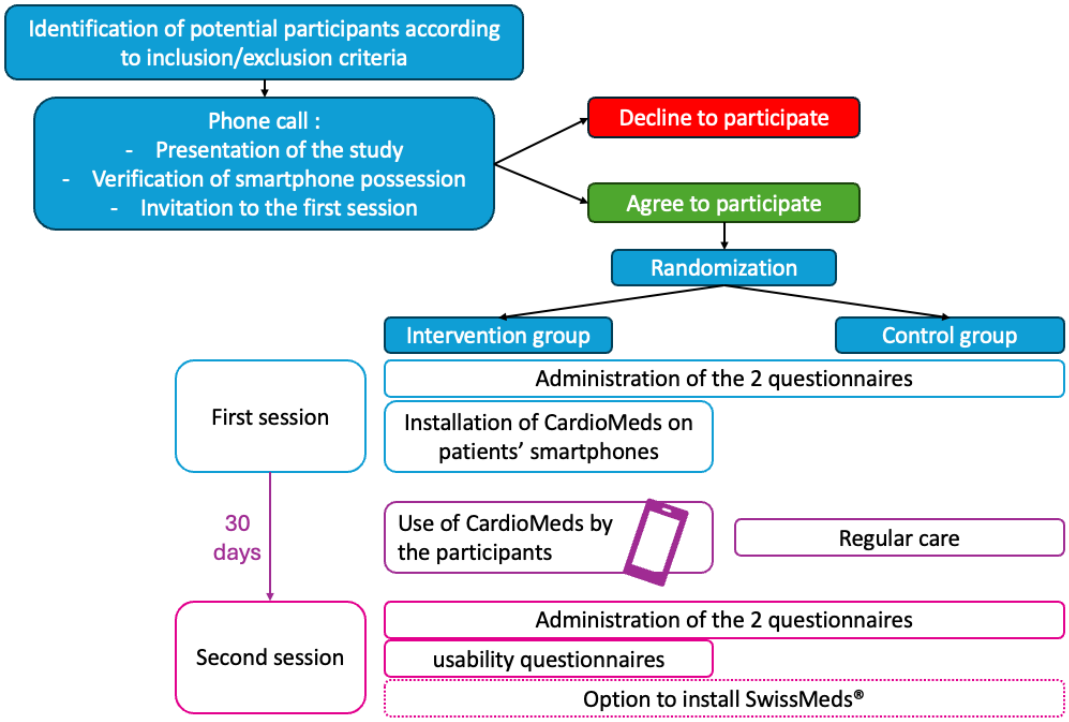
Study Procedures

After providing oral consent and before the first session, participants were randomized using a computer-generated random allocation sequence with a 1:1 ratio between the intervention and control groups, without stratification or blocking. The allocation sequence was generated by a member of the research team not involved in participant recruitment or outcome assessment. Randomization was conducted prior to the first study session for logistical reasons related to group-based study organization. Participants were enrolled by study investigators, and baseline questionnaires were administered before any information regarding group allocation was disclosed (Figure 7). Neither participants nor study personnel responsible for distributing and collecting the questionnaires were aware of group assignment at the time of baseline data collection. Group allocation was revealed only after completion of baseline assessments, at which point participants in the intervention group received instructions for app installation. Given the nature of the digital intervention, blinding of participants and study staff after allocation was not feasible. Outcome assessment relied on self-administered

XSL•FO
RenderX

questionnaires, and data analysis was conducted on deidentified datasets by analysts not involved in intervention delivery.

Figure 7. Flowchart of the study design and procedures of a pilot randomized controlled trial evaluating the Cardio-Meds mobile app. The figure outlines participant recruitment, randomization, intervention, and follow-up procedures in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) between March and November 2024. Adult patients with heart failure were randomized to either a 30-day Cardio-Meds intervention plus usual care or to usual care alone. Heart failure knowledge and medication adherence were assessed at baseline and at 30 days using the Dutch Heart Failure Knowledge Scale (DHFKS) and the Basel Assessment of Adherence to Immunosuppressive Medication Scale (BAASIS).



During session 1, after written informed consent, participants completed 2 baseline assessments:

- The Dutch Heart Failure Knowledge Scale (DHFKS) assessing HF-related knowledge and self-management
- The Basel Assessment of Adherence to Immunosuppressive Medication Scale (BAASIS) evaluating medication adherence

Intervention group participants installed the Cardio-Meds app on their smartphones with guidance on usage and for medication entry. A brief standardized mid-study support call was conducted solely to identify and resolve potential technical issues related to app installation or functionality. No medical advice, educational reinforcement, or adherence-related counseling was provided during this call. After 30 days, all participants returned for a second session, completing the DHFKS and BAASIS again. Intervention participants also completed a usability questionnaire and had the possibility to give nondirected feedback on the app. At the end of the study, both groups could install SwissMeds, the publicly released version of the app, soon to include modules for other chronic conditions.

Most participants attended group sessions (5 - 12 participants) held in March, May, and November 2024. Remaining participants were seen individually during the week following each of these sessions.

The following anonymized usage data were collected from the app:

- Total time spent on the app
- Number of logins
- Time spent on the information section
- Number of weight, blood pressure, and heart rate entries over the 30-day follow-up period
- Percentage of correct answers to the quizzes

Clinical and personal data were collected from the electronic medical record at HUG, anonymized, and securely stored on HUG servers for 10 years in compliance with institutional policy.

Outcome Measures

Primary Outcome

HF knowledge and self-management were assessed using the DHFKS, a validated 15-item questionnaire covering general knowledge, dietary recommendations, sodium and fluid restrictions, and recognition of alarm symptoms. This tool was selected for its validity and sensitivity in evaluating the effect of therapeutic education in patients with HF. The questionnaire was translated into French using neural machine translation (DeepL) and reviewed for clinical accuracy and clarity by an experienced HF specialist (PM). No formal forward-backward translation or psychometric validation of the French version was performed, and the scale was therefore used as an exploratory measure to assess within-group changes in knowledge rather than as a fully validated French-language instrument. Scores range from 0 to 15, with higher scores indicating better knowledge. Based on pragmatic and clinical

XSL•FO
RenderX

considerations, an absolute improvement of 2 points on the DHFKS was prespecified as a meaningful change for sample size estimation. This threshold was used as an assumption for planning purposes rather than as a validated minimal clinically important difference (Multimedia Appendix 1).

Secondary Outcome

Medication adherence was assessed using the BAASIS, a structured self-report questionnaire originally validated in transplant populations. In the absence of an HF-specific validated adherence questionnaire at the time of study design, the BAASIS was used as an exploratory tool to capture core adherence behaviors (initiation, implementation, and persistence) relevant to chronic disease management. BAASIS covers three dimensions of adherence behavior:

- Initiation (starting newly prescribed medications during the reference period)
- Implementation (missed doses, wrong dosage, or timing)
- Persistence (discontinuation without medical advice)

Responses are binary (“yes”/“no”), with frequency details for selected items. For the “Initiation” domain, analyses were restricted to participants who had a newly prescribed medication during the respective reference period, as specified by the BAASIS instrument; therefore, denominators vary across time points and do not reflect missing data or participant exclusion. A validated French version was used (Multimedia Appendix 2).

Exploratory Outcome

Usability of the app was evaluated in the intervention group using the System Usability Scale (SUS), a 10-item questionnaire with responses on a 5-point Likert scale. The final score ranges from 0 to 100, with higher scores reflecting better usability (Multimedia Appendix 3).

Statistical Analysis

Descriptive statistics were used to summarize baseline characteristics and outcomes. Continuous variables are presented as mean (SD) and were compared between groups using Student *t* test when appropriate. Categorical variables are presented as frequencies and percentages and were compared using chi-square or Fisher exact tests, depending on expected cell counts.

The primary outcome, change in HF knowledge, was analyzed using an ANCOVA, with the postintervention DHFKS score as the dependent variable, study group as the independent variable, and baseline DHFKS score as a covariate. No post hoc adjustment was performed for other baseline variables, given the limited sample size and exploratory nature of this pilot study.

Analysis Population

All outcome analyses were conducted according to randomized group assignment. Primary and secondary outcome analyses included all participants who completed both baseline and follow-up assessments (completers-only analysis). Participants with missing follow-up data were not included in the corresponding outcome analyses. No participants were excluded post randomization due to intervention-related reasons; analyses were restricted to participants with available baseline and follow-up data for the corresponding outcomes. Secondary outcomes related to medication adherence, assessed using the BAASIS questionnaire (initiation, implementation, and persistence domains), were analyzed descriptively and compared between groups using chi-square or Fisher exact tests, as appropriate. Usability outcomes measured by the SUS and app usage metrics were analyzed descriptively.

The sample size was estimated to inform feasibility rather than to power a definitive efficacy trial. Based on a mean control group DHFKS score of 10.9 (SD 2.3), an anticipated absolute between-group difference of 2 points, a power of 80%, and a 2-sided significance level (α) of .05, an initial sample size of approximately 40 - 45 participants was estimated. Allowing for an anticipated dropout rate of 25% to 30%, a target sample size of 50 participants (25 per group) was considered feasible and appropriate for this pilot exploratory RCT. The study was powered for the primary knowledge outcome only and was not designed to detect differences in secondary outcomes such as medication adherence, particularly in a population with high baseline adherence.

All analyses were performed using R software (version 4.4.2; R Foundation for Statistical Computing). Statistical significance was defined as a 2-sided $P<.05$.

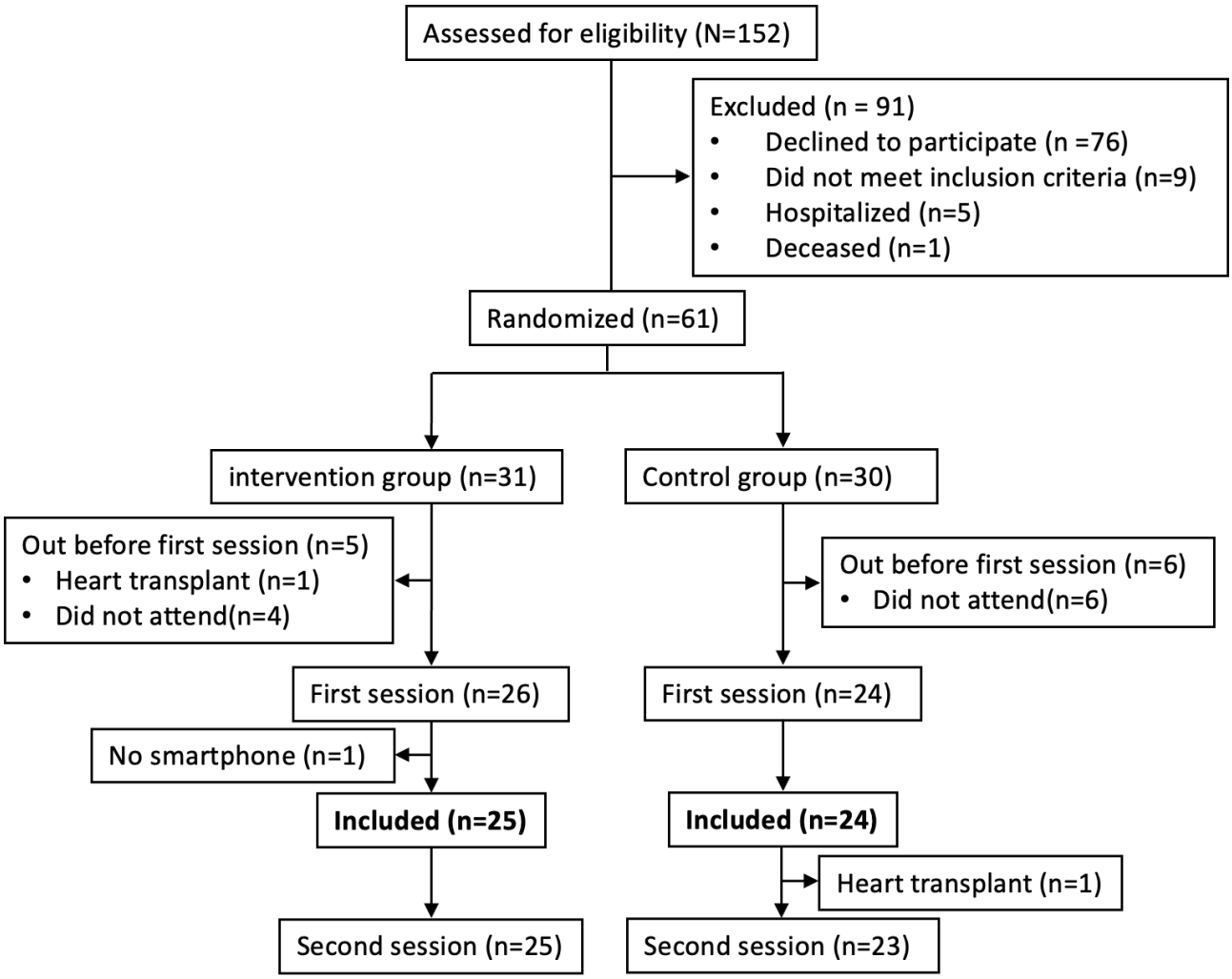
Results

Participants Characteristics

A total of 152 patients with stable HFrEF or HFmrEF followed in the outpatient clinic or undergoing rehabilitation at HUG were preselected by nurses (Figure 8). After the initial phone contact, 61 patients met the eligibility criteria and agreed to participate, while 91 were excluded: 76 declined participation, 9 did not meet the inclusion criteria, 5 were hospitalized, and 1 had died. Among the 61 participants, 31 were randomly assigned to the intervention group and 30 to the control group. Before the first session, 11 participants dropped out: 10 did not attend the scheduled session, and 1 participant in the intervention group underwent heart transplantation. Ultimately, 50 participants attended the first session, and 49 were included, as 1 participant did not own a smartphone. During follow-up, 1 participant in the control group underwent heart transplantation and was therefore unable to attend the second session, and 2 participants did not complete the BAASIS questionnaire at follow-up.

XSL•FO
RenderX

Figure 8. Flowchart of patient recruitment, randomization, and follow-up. The flow diagram shows screening, inclusion, randomization, and follow-up of adult outpatients with heart failure recruited at Geneva University Hospitals (Geneva, Switzerland) between March and November 2024 for a pilot randomized controlled trial of the Cardio-Meds mobile app.



This pilot study included a total of 49 adult participants with HF, of whom 78% (38/49) were male, with a mean age of 62 (SD 11.4) years. Most participants were French- or Italian-speaking and demonstrated a good level of comprehension of French. Both ischemic and nonischemic etiologies of HF were represented, with a mean left ventricular ejection fraction of 37.2% (SD 12). Participants were generally polymorbid, with common comorbidities including diabetes, hypertension, and atherosclerotic cardiovascular disease. All participants were receiving guideline-directed medical therapy, including angiotensin-converting enzyme inhibitor/angiotensin receptor blocker/angiotensin receptor–neprilysin inhibitors, beta-blockers, mineralocorticoid receptor antagonists, and sodium-glucose cotransporter-2 inhibitors. The average number of daily medications and intakes was high (more than 8 medications per day), reflecting the complexity of treatment in this population. Overall, the cohort was representative of a typical population of individuals with HF managed in an outpatient and rehabilitation setting (Table 1).

XSL•FO
RenderX

Table . Baseline characteristics of participants included in a pilot randomized controlled trial of a mobile health intervention for heart failure (N=49).[a]

Characteristic	Intervention group (n=25)[b]	Control group (n=24)[b]	*P* value[c]
Gender, n (%)			.67
Female	5 (20)	6 (25)	
Male	20 (80)	18 (75)	
Mean age (y), mean (SD)	59.2 (10.3)	65.2 (11.9)	.06
HF[d] etiology, n (%)			.98
Ischemic	10 (40)	9 (37.5)	
Arrhythmic	4 (16)	5 (20.8)	
Valvular	3 (12)	2 (8.3)	
Toxic	6 (24)	3 (12.5)	
Genetic	6 (24%)	5 (20.8)	
Hypertensive	5 (20)	4 (16.7)	
Infiltrative	3 (12)	3 (12.5)	
Idiopathic	4 (16)	3 (12.5)	
Other	1 (4)	0 (0)	
Time since diagnosis (mo), mean (SD)	55 (85.6)	66.3 (87.6)	.65
HF medication, n (%)			>.99
ACE[e]/ARB[f]/ARNI[g]	23 (92)	22 (91.7)	
BB[h]	21 (84)	21 (87.5)	
MRA[i]	16 (64)	17 (70.8)	
SGLT2[j]	23 (92)	21 (87.5)	
Loop diuretic	13 (52)	13 (54.2)	
Number of medications, mean (SD)	8.5 (3.6)	8.9 (3.1)	.65
Daily intakes, mean (SD)	9.6 (3.8)	10 (4.3)	.76
Comorbidities, n (%)			.13
AF[k]	8 (32)	8 (33.3)	
COPD[l]	0 (0)	1 (4.2)	
PAD[m]/Stroke	8 (32)	2 (8.3)	
Diabetes	3 (12)	8 (33.3)	
Hypertension	13 (52)	10 (41.7)	
LVEF[n], mean (SD)	38 (12)	37 (12)	.69
NYHA[o], mean (SD)	1.8 (0.9)	1.8 (1.1)	.75
Employment status, n (%)			.56
Full-time	13 (52)	8 (33.3)	
Part-time	3 (12)	3 (12.5)	
Unemployed or disability benefits	2 (8)	2 (8.3)	
Retired	7 (28)	11 (45.8)	
Ethnicity, n (%)			.64
Caucasian	22 (88)	20 (83.3)	
African	3 (12)	4 (16.7)	

XSL•FO
RenderX

Characteristic	Intervention group (n=25)[b]	Control group (n=24)[b]	*P* value[c]
Level of French comprehension and expression, n (%)			.61
Excellent	19 (76)	19 (79.2)	
Good	5 (20)	5 (20.8)	
Average	1 (4)	0 (0)	
Mother tongue, n (%)			.25
French	12 (48)	16 (66.7)	
German	0 (0)	1 (4.2)	
Italian	2 (8)	3 (12.5)	
English	1 (4)	0 (0)	
Other	10 (40)	4 (16.7)	

[a]Baseline demographic, clinical, and treatment characteristics of adult outpatients with heart failure with reduced or mildly reduced ejection fraction enrolled at Geneva University Hospitals (Geneva, Switzerland) between March and November 2024, stratified by randomized group (Cardio-Meds intervention vs usual care).

[b]Baseline characteristics are reported for participants who attended the first study session and completed baseline assessments.

[c]Baseline group comparisons are presented for descriptive purposes only and should not be interpreted as formal tests of randomization balance.

[d]HF: heart failure.

[e]ACE: angiotensin-converting enzyme.

[f]ARB: angiotensin II receptor blocker.

[g]ARNI: angiotensin receptor–neprilysin inhibitor.

[h]BB: beta-blocker.

[i]MRA: mineralocorticoid receptor antagonist.

[j]SGLT2: sodium-glucose cotransporter 2.

[k]AF: atrial fibrillation.

[l]COPD: chronic obstructive pulmonary disease.

[m]PAD: peripheral artery disease.

[n]LVEF: left ventricular ejection fraction.

[o]NYHA: New York Heart Association.

App Usage in Test Group

During the 30-day follow-up period, participants in the intervention group used the app for a mean total duration of 177.4 (SD 130.4) minutes, with a mean of 48.6 (SD 36.2) logins. App usage was highly variable across participants. Total usage time had a median of 123 (IQR 74 - 273) minutes, and the number of logins had a median of 43 (IQR 19 - 85), indicating a skewed distribution with a subset of highly active users. The mean time spent per login was 5.3 (SD 3.6) minutes. Participants spent a mean of 7.9 (SD 9.9) minutes on the information section. Weight was recorded a mean of 5.9 (SD 9.6) times, blood pressure and heart rate were entered a mean of 6.8 (SD 10.3) times, corresponding to an average of approximately 1 entry every 4 to 5 days. Among participants who accessed the quiz feature, a mean of 16.4 (SD 17.6) questions were answered, with a mean of 12.5 (SD 13.9) correct answers, corresponding to a mean accuracy rate of 0.7 (SD 0.3) (Multimedia Appendix 4).

HF Knowledge and Self-Management (Primary Outcome)

The DHFKS questionnaire was administered at both study sessions. At baseline (session 1), the mean score was 11.1 (SD 2.4) in the intervention group and 10.5 (SD 2.9) in the control group. In the primary analysis using ANCOVA adjusted for baseline DHFKS score, the intervention group had significantly higher follow-up knowledge scores than the control group, with an adjusted mean between-group difference of +1.3 points (95% CI 0.40 - 2.20; $P<.001$). For descriptive purposes, within-group analyses showed that DHFKS scores remained stable in the control group (10.5 [SD 2.9] at baseline vs 10.4 [SD 3] at follow-up; $P=.82$), whereas the intervention group demonstrated a statistically significant increase (11.1 [SD 2.4] to 12.4 [SD 2.4]; $P<.001$) (Figure 9)(Table 2). Although the improvement in the intervention group was statistically significant (+1.3 points), it did not reach the prespecified +2-point threshold used for sample size estimation. No significant association was observed between app usage metrics (including total usage time and number of logins) and change in DHFKS score .

XSL•FO
RenderX

Figure 9. Heart failure knowledge scores before and after a 30-day mobile health intervention. Boxplots show Dutch Heart Failure Knowledge Scale (DHFKS) scores at baseline and after 30 days in the intervention group (Cardio-Meds plus usual care) and the control group (usual care alone). The data are derived from a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) in adult patients with heart failure between March and November 2024.

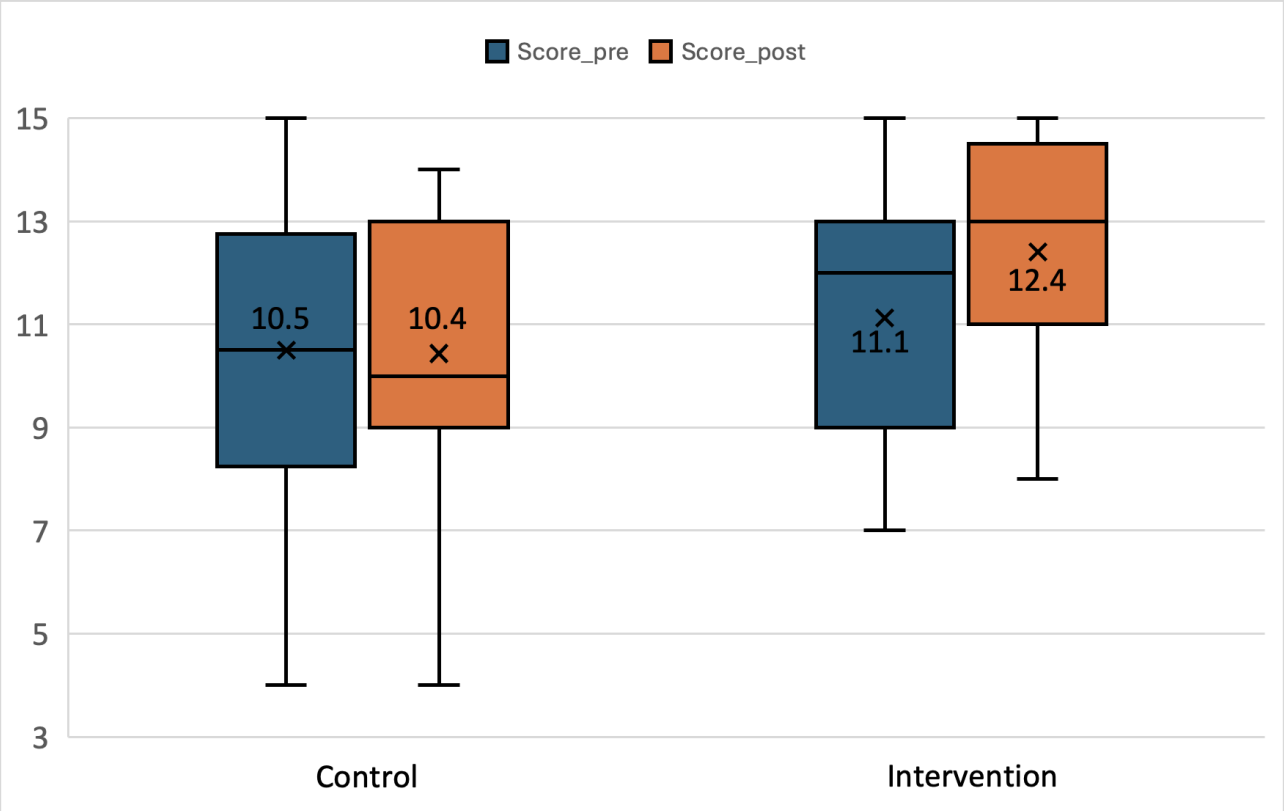


Table . Heart failure knowledge before and after a 30-day intervention using the Cardio-Meds mobile app.[a]

Group	Preintervention, mean (SD)	Postintervention, mean (SD)	Mean change	*P* value
Control	10.5 (2.9)	10.4 (3)	–0.1	.82[b]
Intervention	11.1 (2.4)	12.4 (2.4)	1.3	<.001[b]
Adjusted between-group difference (ANCOVA[c])	—[d]	—[d]	1.3 (95% CI 0.40-2.20)	<.001[e]

[a]Dutch Heart Failure Knowledge Scale scores at baseline and after 30 days are shown for intervention and control groups in a single-center pilot randomized controlled trial conducted at Geneva University Hospitals (Geneva, Switzerland) among adult patients with heart failure in 2024. Within-group changes and adjusted between-group differences are reported.

[b]Paired *t* test (within-group).

[c]ANCOVA: analysis of covariance.

[d]Not applicable.

[e]ANCOVA adjusted for baseline DHFKS score.

Medication Adherence (Secondary Outcome) With BAASIS

For the initiation domain, analyses were restricted to participants with newly prescribed medications during the reference period, resulting in very small sample sizes (control group: n=5 at follow-up; intervention group: n=7). The intervention group maintained a score of 100% at both time points, while the control group showed a decrease from 100% at baseline to 80% at 1 month. The between-group comparison was not statistically significant (*P*=.38); however, this result should be interpreted cautiously given the very limited sample size and lack of statistical power for this specific domain. As per the BAASIS instrument structure, denominators for the "Initiation" domain vary across assessment time points depending on whether participants had newly prescribed medications during the reference period. For the implementation domain, the intervention group showed a decrease from 64% at baseline to 52% at 1 month. In contrast, the control group showed a slight increase, from 46% to 48%, likely due to missing responses from three participants at follow-up. The between-group difference was not statistically significant (*P*=.61). For the

XSL•FO
RenderX

persistence domain, the intervention group decreased slightly from 100% to 96% after the intervention, while the control group remained relatively stable, from 96% to 95% (Table 3). These small variations correspond to 1 participant in each group: in the control group, 1 answered "yes" at both time points; in the intervention group, 1 answered "yes" only at the second session. The between-group comparison showed no statistically significant difference (P=.76).

Table . Medication adherence outcomes assessed with the BAASIS questionnaire before and after a 30-day follow-up period.[a]

Domain and group	Preintervention, n/N (%)	Postintervention, n/N (%)	*P* value
Initiation			.38
Control	4/4 (100)	4/5 (80)	
Intervention	14/14 (100)	7/7 (100)	
Implementation			.61
Control	11/24 (46)	10/21 (48)	
Intervention	16/25 (64)	13/25 (52)	
Persistence			.76
Control	23/24 (96)	20/21 (95)	
Intervention	25/25 (100)	24/25 (96)	

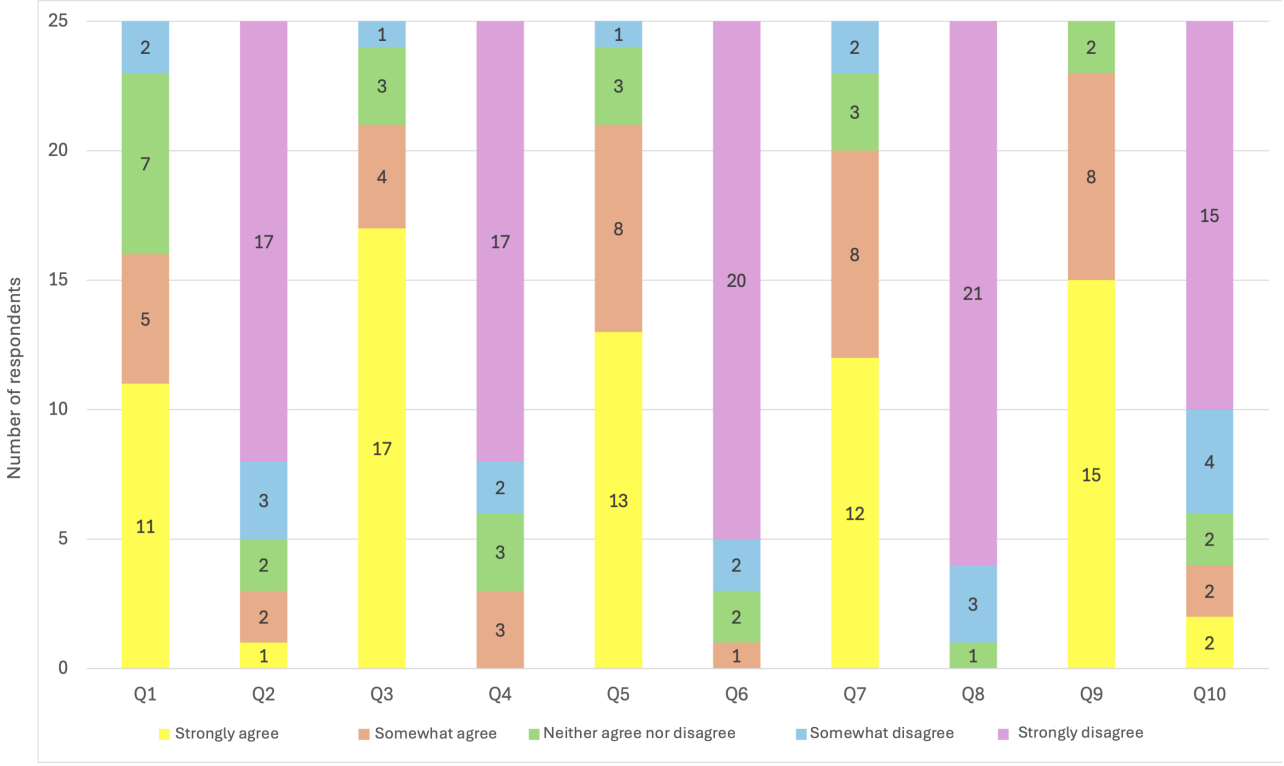
[a]Initiation, implementation, and persistence domains of the Basel Assessment of Adherence to Immunosuppressive Medication Scale (BAASIS) at baseline and 30 days in intervention and control groups. Data derived from a pilot randomized controlled trial evaluating a mobile health intervention in adult outpatients with heart failure conducted at Geneva University Hospitals (Geneva, Switzerland) between March and November 2024. Initiation denominators reflect only participants with newly prescribed medications during the reference period, according to BAASIS specifications.

Usability of the App (Exploratory Outcome)

The usability of the app was assessed using the SUS, a standardized questionnaire designed to evaluate user experience and ease of use. The mean SUS score in the intervention group was 84.3 (SD 15) (Multimedia Appendix 5), indicating excellent usability, as it falls within the 90th percentile of usability scores [24]. Notably, 64% (16/25) of participants indicated they would continue using the app (Figure 10). However, beyond the predefined usage metrics collected (eg, total time, logins, section time, and entries), no additional backend analytics were available to characterize navigation pathways or feature-level sequences. Future studies should include backend analytics to better understand interaction dynamics with the app. Across all SUS items, most of the responses reflected positive user experience and good perceived usability of the Cardio-Meds app (Figure 10).

XSL•FO
RenderX

Figure 10. System Usability Scale (SUS) item responses for the Cardio-Meds mobile app. Likert-style plots display responses to the 10 items of the SUS completed after 30 days of app use by participants randomized to the intervention group in a pilot randomized controlled trial. The study was conducted at Geneva University Hospitals (Geneva, Switzerland) between March and November 2024 among adult outpatients with heart failure.



Participant Feedback

Participants could provide comments at the end of the questionnaires. Overall, the app was seen as well-designed and pleasant to use, with 1 participant describing it as a "companion that sets the rhythm of his day." Users appreciated features such as the motivational message "Everything is fine today," medication images for easy identification, and the reminder function. Quizzes and educational content were well received. Reported difficulties included deleting medications after input errors, handling missed doses, and grouping medications, often due to unfamiliarity with the app's functions. Suggestions included improving text contrast, syncing reminders with smartwatches, managing time zones, integrating with health platforms, storing prescriptions, and adding tracking for weight and blood pressure. Some participants also requested video content and information on disease progression. These comments will inform future app development.

Discussion

Principal Findings

In this pilot RCT conducted in a stable ambulatory HF population, a 30-day intervention using the Cardio-Meds mobile app was associated with a statistically significant short-term improvement in HF knowledge and self-management compared with usual care. After adjustment for baseline knowledge, the intervention group demonstrated a statistically significant between-group improvement of 1.3 points in the DHFKS score compared with the control group. The magnitude of improvement, while statistically significant, was smaller than the prespecified effect size and should therefore be interpreted as an exploratory signal rather than a clinically definitive change. In contrast, no significant differences were observed between groups in self-reported medication adherence across the BAASIS domains; small numerical variations should be interpreted cautiously in light of the pilot design and reliance on self-reported measures. The app demonstrated high usability and acceptability, with excellent SUS scores and a majority of participants reporting willingness to continue using the app. Taken together, these findings should be interpreted as short-term, exploratory results primarily informing feasibility, usability, and educational impact rather than durable behavioral change.

Comparison With Prior Work

Previous digital health interventions for HF have reported mixed but generally encouraging effects on self-care behaviors and symptom outcomes. For example, Yoon et al [14] found that a smartphone-based intervention with remote feedback improved HF symptom burden over follow-up, and Kitsiou et al [13] showed that multiapp interventions may enhance self-care engagement. Systematic evidence also supports the value of digital health tools in HF management, reinforcing the need for continued innovation and larger, long-term studies [15]. In this context, the improvement in HF knowledge observed in our study is consistent with prior evidence showing that digital interventions can enhance patient understanding and engagement. A recent systematic review by Mouselimis et al [25] reported that mHealth tools frequently improve disease knowledge and engagement, although effects on clinical outcomes and adherence remain heterogeneous. Fernández-Gutiérrez et al [20] emphasized the importance of validated, patient-centered educational content in HF apps, an

XSL•FO
RenderX

aspect directly addressed in Cardio-Meds through guideline-based information and interactive quizzes. Similarly, Choi et al [26] demonstrated that an HF-specific mobile app improved functional and echocardiographic parameters, although such clinical endpoints were not assessed in the present study. Unlike the findings of Ruppar et al [10], who reported improvements in medication adherence following adherence-focused interventions, we did not observe a significant adherence effect. This discrepancy may partly be explained by the characteristics of the study population. Baseline adherence was high in the persistence domain, with nearly all participants reporting continued use of prescribed medications, reflecting a stable ambulatory population with established treatment routines. In contrast, baseline adherence in the implementation domain was suboptimal in both groups, indicating difficulties related to day-to-day medication-taking behavior rather than treatment discontinuation. Despite this, no significant improvement in implementation adherence was observed after the 30-day intervention, possibly due to the short follow-up period, reliance on self-reported measures, and the exploratory nature of the study. Usability and acceptability are critical determinants of digital intervention success, and Cardio-Meds performed well in this regard, with a mean SUS score of 84, consistent with excellent usability [24] and comparable to other successful HF apps.

Limitations

Several limitations should be acknowledged. First, this pilot study was powered exclusively for the primary outcome of HF knowledge. Consequently, it was underpowered to detect differences in secondary outcomes such as medication adherence, particularly in a stable ambulatory population with high baseline adherence. Adherence findings should therefore be interpreted as exploratory and hypothesis-generating rather than confirmatory.

Second, the short follow-up period limited assessment of longer-term knowledge retention, sustained behavioral change, and clinical outcomes. Short-term improvements may partly reflect transient or “honeymoon” effects, and longer follow-up will be required to evaluate the durability of effects on self-management and adherence.

Third, medication adherence was assessed using self-reported measures, which are inherently subject to recall and social desirability bias. Although validated instruments were used, future studies would benefit from incorporating objective adherence measures such as pharmacy refill data or electronic monitoring. In this context, the small, non-significant decrease in implementation adherence observed in the intervention group may reflect increased awareness of medication-taking behavior and more accurate self-reporting following exposure to the intervention and repeated questioning, rather than a true deterioration in adherence.

Fourth, the single-center design and inclusion of a stable, chronically treated outpatient population may limit generalizability, particularly to recently hospitalized or higher-risk patients.

Fifth, HF knowledge was assessed using a French translation of the DHFKS that did not undergo formal linguistic or psychometric validation. While the items are factual and clinically straightforward and the original scale was developed in a population comparable to the Swiss context, this limits the interpretation of absolute score values and the strength of conclusions regarding knowledge gains. In addition, the absence of a validated minimal clinically important difference for the DHFKS constrains assessment of the clinical relevance of the observed improvement. Accordingly, knowledge findings should be considered exploratory.

Sixth, app engagement showed marked heterogeneity, with a small number of highly active users and overall limited use of specific features such as physiological self-monitoring. This skewed distribution limits the interpretation of average usage statistics and precluded meaningful dose-response analyses. Consistent with this, no clear association was observed between engagement metrics and improvement in knowledge scores. Although a testing effect from repeated questionnaire administration may have contributed to knowledge gains, this would be expected to affect both groups similarly, and the absence of improvement in the control group suggests that testing alone is unlikely to explain the observed between-group difference.

Finally, the intervention group received a brief mid-study technical support call that was not mirrored in the control group. Although this contact was limited to troubleshooting and did not include educational or behavioral guidance, an attention-related or Hawthorne effect cannot be fully excluded. In addition, baseline imbalances were observed between groups, with the intervention group being younger and including more non-native French speakers. Such differences may occur by chance in small pilot studies and could have influenced engagement, as age may proxy digital literacy. Given the limited sample size, post hoc adjustment was not performed, and these factors should be considered potential confounders.

Future Directions

Future research should evaluate Cardio-Meds in larger, multicenter randomized trials with longer follow-up and inclusion of clinical endpoints such as hospitalization, emergency visits, or quality of life. Studying the intervention in the early post-discharge period, when patients are particularly vulnerable and educational needs are high, may be especially informative. Further development could also include adaptation to different literacy levels and languages to enhance accessibility and equity of care.

Conclusions

In this pilot RCT, use of the Cardio-Meds mobile app was associated with a statistically significant but modest short-term improvement in HF knowledge compared with usual care, without measurable changes in self-reported medication adherence. The app demonstrated high usability and acceptability in a stable ambulatory HF population. These findings should be interpreted as exploratory and hypothesis-generating and support further evaluation of

XSL•FO
RenderX

Cardio-Meds in larger studies with longer follow-up and clinically relevant outcomes.

Acknowledgments

The authors would like to thank the patients who participated in this study, as well as the heart failure nurses and staff of the Cardiology Division and Cardiac Rehabilitation Program at Geneva University Hospitals for their support with patient recruitment and study logistics. We also acknowledge the contributions of the Information Technology and Pharmacy departments involved in the development and maintenance of the Cardio-Meds app.

The authors confirm that all data reported in this manuscript are original, were collected at Geneva University Hospitals, and were analyzed by the study team. Generative artificial intelligence (ChatGPT; OpenAI) was used solely to assist with language editing and drafting of selected sections of the manuscript and response-to-reviewers documents under full author supervision. In addition, DeepL was used to support translation of study materials, as described in the Methods. Neither tool was used for data collection, statistical analyses, outcome assessment, or generation of results. All content was critically reviewed, verified, and approved by the authors, who take full responsibility for the accuracy, integrity, and scientific content of the work.

Funding

This study was supported by the GEcor Foundation (Fondation privée de la cardiologie universitaire à Genève). The funding body had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the manuscript for publication.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to institutional data protection policies and the inclusion of sensitive health-related information. Deidentified data may be made available from the corresponding author upon reasonable request and subject to approval by Geneva University Hospitals, in accordance with applicable regulations.

Authors' Contributions

Conceptualization: PM, FE, KB, EM
Data curation: VB, HH
Formal analysis: VB, PM
Investigation: VB, ET, LS, GG, AS-P, SP, LG
Methodology: PM, FE, KB, EM
Project administration: PM
Software: FE, HH
Supervision: PM
Writing – original draft: VB, PM
Writing – review & editing: All authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dutch Heart Failure Knowledge Scale questionnaire translated in French with DeepL.
[DOCX File, 17 KB - cardio_v10i1e83022_app1.docx]

Multimedia Appendix 2
Basel Assessment of Adherence to Immunosuppressive Medications Scale (BAASIS) in French.
[DOCX File, 27 KB - cardio_v10i1e83022_app2.docx]

Multimedia Appendix 3
Usability questionnaire.
[DOCX File, 285 KB - cardio_v10i1e83022_app3.docx]

Multimedia Appendix 4
Usage of the app in the intervention group during the 30 days follow-up.
[DOCX File, 19 KB - cardio_v10i1e83022_app4.docx]

XSL•FO
RenderX

Multimedia Appendix 5
System Usability Scale (SUS) questionnaire responses.
[DOCX File, 19 KB - cardio_v10i1e83022_app5.docx]

Checklist 1
CONSORT checklist.
[PDF File, 465 KB - cardio_v10i1e83022_app6.pdf]

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
2. Vivaldo M, Alexe R, Sierro C, Tapponnier M, Djokic B, Girod G. Insuffisance cardiaque en 2021: un tour d'horizon pour le généraliste [Article in French]. Rev Med Suisse 2021;17(728):424-428. [doi: 10.53738/REVMED.2021.17.728.0424] [Medline: 33656294]
3. Crespo-Leiro MG, Anker SD, Maggioni AP, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT): 1-year follow-up outcomes and differences across regions. Eur J Heart Fail 2016 Jun;18(6):613-625. [doi: 10.1002/ejhf.566] [Medline: 27324686]
4. Platz E, Jhund PS, Claggett BL, et al. Prevalence and prognostic importance of precipitating factors leading to heart failure hospitalization: recurrent hospitalizations and mortality. Eur J Heart Fail 2018 Feb;20(2):295-303. [doi: 10.1002/ejhf.901] [Medline: 28872259]
5. Lesyuk W, Kriza C, Kolominsky-Rabas P. Cost-of-illness studies in heart failure: a systematic review 2004-2016. BMC Cardiovasc Disord 2018 May 2;18(1):74. [doi: 10.1186/s12872-018-0815-3] [Medline: 29716540]
6. McDonagh TA, Metra M, Adamo M, et al. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2023 Oct 1;44(37):3627-3639. [doi: 10.1093/eurheartj/ehad195] [Medline: 37622666]
7. Taylor RS, Dalal HM, Zwisler AD. Cardiac rehabilitation for heart failure: 'Cinderella' or evidence-based pillar of care? Eur Heart J 2023 May 1;44(17):1511-1518. [doi: 10.1093/eurheartj/ehad118] [Medline: 36905176]
8. Van Spall HGC, Rahman T, Mytton O, et al. Comparative effectiveness of transitional care services in patients discharged from the hospital with heart failure: a systematic review and network meta-analysis. Eur J Heart Fail 2017 Nov;19(11):1427-1443. [doi: 10.1002/ejhf.765] [Medline: 28233442]
9. Son YJ, Choi J, Lee HJ. Effectiveness of nurse-led heart failure self-care education on health outcomes of heart failure patients: a systematic review and meta-analysis. Int J Environ Res Public Health 2020 Sep 9;17(18):6559. [doi: 10.3390/ijerph17186559] [Medline: 32916907]
10. Ruppar TM, Cooper PS, Mehr DR, Delgado JM, Dunbar-Jacob JM. Medication adherence interventions improve heart failure mortality and readmission rates: systematic review and meta-analysis of controlled trials. J Am Heart Assoc 2016 Jun 17;5(6):e002606. [doi: 10.1161/JAHA.115.002606] [Medline: 27317347]
11. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021 Sep 21;42(36):3599-3726. [doi: 10.1093/eurheartj/ehab368] [Medline: 34447992]
12. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 7;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
13. Kitsiou S, Gerber BS, Buchholz SW, Kansal MM, Sun J, Pressler SJ. Patient-centered mHealth intervention to improve self-care in patients with chronic heart failure: phase 1 randomized controlled trial. J Med Internet Res 2025 Jan 15;27:e55586. [doi: 10.2196/55586] [Medline: 39813671]
14. Yoon M, Lee S, Choi JY, et al. Effectiveness of a smartphone app–based intervention with Bluetooth-connected monitoring devices and a feedback system in heart failure (SMART-HF Trial): randomized controlled trial. J Med Internet Res 2024 Apr 29;26:e52075. [doi: 10.2196/52075] [Medline: 38683665]
15. Azizi Z, Broadwin C, Islam S, et al. Digital health interventions for heart failure management in underserved rural areas of the United States: a systematic review of randomized trials. J Am Heart Assoc 2024 Jan 16;13(2):e030956. [doi: 10.1161/JAHA.123.030956] [Medline: 38226517]
16. Dunn Lopez K, Chae S, Michele G, et al. Improved readability and functions needed for mHealth apps targeting patients with heart failure: an app store review. Res Nurs Health 2021 Feb;44(1):71-80. [doi: 10.1002/nur.22078] [Medline: 33107056]
17. WOW ME 2000mg - heart failure self-management tool for patients and caregivers. Download CNET. URL: https://download.cnet.com/wow-me-2000mg-heart-failure-self-management-tool-for-patients-and-caregivers/3000-2129_4-76773751.html [accessed 2026-01-30]
18. WebMD: symptom checker. Google Play. URL: https://play.google.com/store/apps/details?id=com.webmd.android [accessed 2026-02-11]

XSL•FO
RenderX

19. 8 smart mobile apps and gadgets for heart health. Preventive Cardiovascular Nurses Association (PCNA). 2025 Apr 15. URL: https://pcna.net/8-smart-mobile-apps-and-gadgets-for-heart-health/ [accessed 2026-01-30]
20. Fernández-Gutiérrez M, Bas-Sarmiento P, Jesús Marín-Paz A, et al. Self-management in heart failure using mHealth: a content validation. Int J Med Inform 2023 Mar;171:104986. [doi: 10.1016/j.ijmedinf.2023.104986] [Medline: 36638582]
21. Bas-Sarmiento P, Fernández-Gutiérrez M, Poza-Méndez M, et al. Needs of patients with multi-morbidity and heart failure for the development of a mHealth to improve their self-management: a qualitative analysis. Digit Health 2023;9:20552076231180466. [doi: 10.1177/20552076231180466] [Medline: 37325072]
22. Simioni L, Tessitore E, Hagberg H, et al. Cardiomeds, an mHealth app for self-management to support Swiss patients with heart failure: 2-stage mixed methods usability study. JMIR Form Res 2025 Jan 15;9:e63941. [doi: 10.2196/63941] [Medline: 39813081]
23. Ehrler F, Gschwind L, Meyer P, Blondon K. Swiss-Meds: an app fostering medication adherence of Swiss patient. Stud Health Technol Inform 2019;259:71-76. [Medline: 30923276]
24. Bangor A, Kortum P, Miller JT. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123 [FREE Full text]
25. Mouselimis D, Tsarouchas A, Vassilikos VP, et al. The role of patient-oriented mHealth interventions in improving heart failure outcomes: a systematic review of the literature. Hellenic J Cardiol 2024;77:81-92. [doi: 10.1016/j.hjc.2023.11.001] [Medline: 37926237]
26. Choi EY, Park JS, Min D, Ahn S, Ahn JA. Heart failure-smart life: a randomized controlled trial of a mobile app for self-management in patients with heart failure. BMC Cardiovasc Disord 2023 Jan 9;23(1):7. [doi: 10.1186/s12872-023-03039-8] [Medline: 36624373]

Abbreviations

BAASIS: Basel Assessment of Adherence to Immunosuppressive Medication Scale
DHFKS: Dutch Heart Failure Knowledge Scale
HF: heart failure
HFmrEF: heart failure with mildly reduced ejection fraction
HFrEF: heart failure with reduced ejection fraction
HUG: Geneva University Hospitals
mHealth: mobile health
RCT: randomized controlled trial
SUS: System Usability Scale

Edited by A Coristine; submitted 26.Aug.2025; peer-reviewed by C Fairhurst, J Santos; revised version received 22.Jan.2026; accepted 23.Jan.2026; published 23.Feb.2026.

Please cite as:
Buswell V, Massie E, Tessitore E, Simioni L, Guebey G, Hagberg H, Schneider-Paccot A, Pant S, Blondon K, Gschwind L, Ehrler F, Meyer P
Impact of the Cardio-Meds Mobile App on Heart Failure Knowledge and Medication Adherence: Pilot Randomized Controlled Trial
JMIR Cardio 2026;10:e83022
URL: https://cardio.jmir.org/2026/1/e83022
doi:10.2196/83022

© Victor Buswell, Emmanuelle Massie, Elena Tessitore, Lisa Simioni, Guillaume Guebey, Hamdi Hagberg, Aurélie Schneider-Paccot, Samaksha Pant, Katherine Blondon, Liliane Gschwind, Frederic Ehrler, Philippe Meyer. Originally published in JMIR Cardio (https://cardio.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Generation of 12-Lead Electrocardiogram Using Smartwatches in Patients With Cardiac Conditions: Clinical Validation Study

Mauro Buelga Suárez[1,2,3], MSN; Carmen Plaza-Seco[4], PhD; Roberto Matía Francés[1,3], MD, PhD; Gonzalo Luis Alonso Salinas[5,6,7], MD, PhD; Manuel Blanco-Velasco[4], PhD

[1]Cardiology Department, Hospital Universitario Ramón y Cajal, Madrid, Spain

[2]Medicine and Health Sciences Department, University of Alcalá, Alcalá de Henares, Madrid, Spain

[3]Ramón y Cajal Institute for Health Research (IRYCIS), Madrid, Spain

[4]Department of Signal Theory and Communications, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain

[5]Department of Cardiology, Faculty of Health Sciences, Universidad Publica de Navarra, Irunlarrea 3, Pamplona, Spain

[6]Health Sciences Department, Universidad Pública de Navarra, Pamplona, Spain

[7]Navarrabiomed (Miguel Servet Foundation), Instituto de Investigación Sanitaria de Navarra (IdiSNA), Pamplona, Spain

Corresponding Author:
Gonzalo Luis Alonso Salinas, MD, PhD
Department of Cardiology, Faculty of Health Sciences, Universidad Publica de Navarra, Irunlarrea 3, Pamplona, Spain

Abstract

Background: Smartwatches have significantly expanded personal health monitoring capabilities; however, their application in comprehensive cardiac assessment remains limited by single-lead electrocardiogram (ECG) acquisition. The inability to directly obtain augmented limb leads represents a major limitation for the evaluation of ischemic and arrhythmogenic conditions.

Objective: This study aimed to present and preliminarily evaluate a methodology for generating augmented limb leads (augmented voltage right [aVR], augmented voltage left [aVL], and augmented voltage foot [aVF]) from sequential smartwatch recordings through signal synchronization and RR interval alignment, and to combine this approach with precordial smartwatch recordings to reconstruct a full 12-lead ECG in a proof-of-concept setting.

Methods: A signal-processing framework was developed to preserve ECG morphology across sequential smartwatch lead recordings. The method is based on RR interval–guided temporal resampling, enabling interlead synchronization and analytical reconstruction of augmented limb leads. A prospective proof-of-concept clinical validation was conducted in 20 cases with diverse cardiac pathologies, including 1 healthy control. Smartwatch-derived ECGs were compared with standard 12-lead ECGs. Evaluation included expert-based clinical interpretation by 85 health care professionals and objective quantitative analysis of key electrocardiographic parameters, including heart rate, PR interval, QRS duration, QT interval, QTc, RR interval, P-wave duration, and PP interval. Mean values, SDs, and intermethod differences were calculated to assess agreement and variability.

Results: Reconstruction of a 12-lead ECG from smartwatch-acquired signals was successfully achieved in all cases. Visual correlation scores between smartwatch-derived and conventional ECGs ranged from 78.1% to 91.3% across specialty-level means (mean 81.1, SD 22.7), indicating high perceived concordance; individual case-level scores ranged more widely, from 47.0% to 96.3%. Evaluator-perceived axis agreement reached 95.4% across all cases, while the investigators' direct axis measurements showed complete concordance in all cases. Quantitative analysis demonstrated small mean differences between modalities, with particularly high agreement observed in structurally stable parameters such as heart rate (mean 77.2, SD 22.3 vs mean 75.9, SD 23.1 bpm; Δ=1.3 bpm) and PR interval (mean 176.2, SD 29.1 vs mean 180.8, SD 19.8 ms; Δ=−4.6 ms). Greater variability was observed in atrial activation parameters such as P-wave duration (Δ=−13.3 ms) and time-dependent parameters such as QT (Δ=−19.2 ms) and QTc (Δ=−12.7 ms).

Conclusions: This proof-of-concept study demonstrates the feasibility of reconstructing a full 12-lead ECG from smartwatch-derived signals using a synchronization-based approach. The results demonstrated generally small mean differences between conventional and smartwatch-derived ECG measurements, although larger interval differences were observed in some individual cases. Importantly, due to the sequential nature of signal acquisition, the proposed methodology does not aim to achieve direct waveform equivalence but rather clinically meaningful agreement in electrocardiographic parameters. These findings support the potential of smartwatches as accessible tools for advanced cardiac monitoring and broaden their applicability beyond single-lead screening.

(JMIR Cardio 2026;10:e86621) doi:10.2196/86621

XSL•FO
RenderX

KEYWORDS

electrocardiography; ambulatory monitoring; telediagnostics; augmented leads; smartwatch

Introduction

In recent decades, rapid advances in technology have enabled the integration of advanced biometric sensors into everyday devices, such as smartwatches. Among the most notable innovations is the incorporation of electrocardiogram (ECG) sensors [1] capable of monitoring the electrical activity of the heart in real time.

Previously limited to functions such as monitoring heart rate and recording physical activity, these devices now offer an accessible tool for detecting potentially serious cardiac abnormalities such as atrial fibrillation (AF) [2-5]. Similarly, some studies have shown that variations in ECG patterns recorded by smartwatches, although limited in their diagnostic capability compared with high-precision medical equipment, may be useful in alerting users to possible cardiac abnormalities. Some devices have detected subtle changes in ST-segment or T-wave irregularities, which, in medical contexts, are considered classic indicators of ischemia and may help identify early ischemic episodes [6,7] and contribute to the detection of ventricular tachycardia (VT) [8,9].

Smartwatches are equipped with specialized software and algorithms that generate electrocardiographic waveforms comparable to those of a single lead (lead I) in a traditional ECG. By applying the basic principles of electrocardiography to smartwatches, it is possible to obtain 9 ECG leads. In the frontal plane, with these devices, only the limb leads (I, II, and III) can be obtained directly. The augmented leads (augmented voltage right [aVR], augmented voltage foot [aVF], and augmented voltage left [aVL]) cannot be calculated directly due to the inability to generate the central terminal (Wilson Central Terminal [WCT]) [10]. This inability to directly derive augmented leads poses a significant diagnostic limitation for cardiac pathologies of arrhythmogenic and ischemic origin, particularly in cases such as inferior infarction (II, III, aVF) or those with a high-lateral location (I, aVL). The main objective of this study is to present and preliminarily evaluate the use of smartwatches for obtaining 12-lead ECGs by generating augmented leads (aVR, aVL, and aVF) from sequentially recorded smartwatch signals. This process uses advanced techniques, including R peak alignment, interpolation, filtering, and signal synchronization. Unlike previous work that averaged beats into a single representative complex [11,12], that compared sequential smartwatch leads without interlead synchronization [13], or that reconstructed missing leads using learning-based generative models [14], our approach uniquely preserves beat-level morphology by aligning RR intervals across sequential recordings and analytically deriving the augmented leads. As secondary objectives, the following were established:

1. Integration of this technique aligns the frontal plane leads with the horizontal plane leads (V1-V6), enabling the reconstruction of a fully synchronized 12-lead ECG.
2. Assessment of the performance in 20 real cases. This methodological validation focused on signal acquisition and processing rather than patient outcomes, with a single institution deliberately chosen to ensure homogeneous conditions and procedural control.

A European patent for this work has been filed under reference EP25382712.5.

Methods

Data Acquisition

In a standard ECG, 12 different leads are recorded simultaneously, each of which provides a detailed view of the electrical activity of the heart from various angles. This simultaneous approach allows for a comprehensive analysis of cardiac electrical behavior. There are 6 leads in the frontal plane, consisting of 3 limb leads (I, II, and III) and 3 augmented limb leads (aVR, aVF, and aVL). The limb leads are the original leads chosen by Einthoven in 1901 [15] to record electrical potentials in the frontal plane. Electrodes are placed on the left arm, right arm, and left leg. These leads are obtained from the difference in electrical potential between each electrode placed on one of the three extremities and the WCT, which is essentially the barycenter of Einthoven's equilateral triangle with a potential very close to infinity (zero). Limb leads are bipolar because they represent an electrical potential difference between 2 poles, positive and negative, with the lead axis being the line connecting the 2 poles (each line is divided into 2 halves, one positive half near the positive pole and one negative half near the negative one). The 3 limb leads are:

- Lead I: potential difference between the left arm (LA) and right arm (RA) (LA − RA).
- Lead II: potential difference between the left leg (LL) and right arm (LL − RA).
- Lead III: potential difference between the left leg and the left arm (LL − LA).

In the horizontal plane, there are 6 leads, called precordial leads (V1, V2, V3, V4, V5, and V6), that measure the electrical activity of the heart at right angles to the frontal plane. They are obtained by placing the electrodes in preset areas of the anterior chest wall. A precordial lead measures the electrical potential difference between a thoracic electrode and the WCT.

When these principles of electrocardiography are used as the basis for a smartwatch, only the limb leads (I, II, and III) can be obtained directly. The augmented leads (aVR, aVF, and aVL) cannot be calculated directly because of the inability to generate the WCT. The polarity of the leads is determined by electrode placement on smartwatches [16].

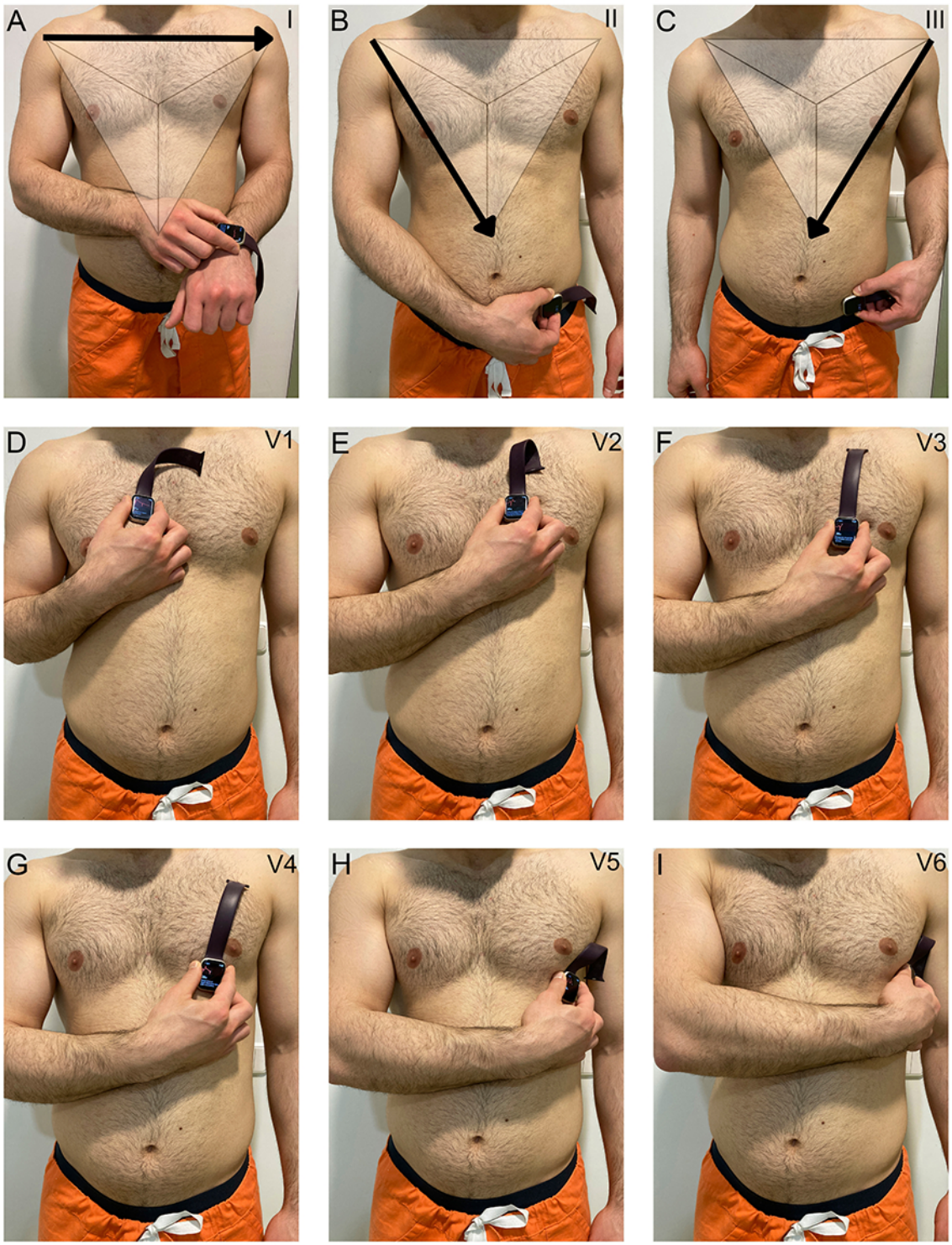
To obtain them, the following approach, shown in Figure 1, is used:

1. The ECG recording of the smartwatch is obtained using its integrated electrodes. The digital crown (negative) and the back crystal (positive) serve as electrodes.

XSL•FO
RenderX

2. To obtain lead I, the watch is placed on the left wrist (back crystal, positive electrode), and the right index finger touches the digital crown (negative electrode).
3. Leads II and III are obtained with the back of the watch (back crystal, positive electrode) against the lower-left abdomen and the right index finger (lead II) or left index finger (lead III) touching the crown (negative electrode).
4. To obtain precordial leads with the smartwatch, the back of the watch (back crystal, positive electrode) is placed in the classic positions for the precordial leads V1-V6, with the right index finger touching the crown (negative electrode).

Figure 1. Schematic of smartwatch placement for obtaining leads I, II, and III (A-C) in the frontal plane and V1 to V6 (D-I) in the horizontal plane. Lead I: potential difference between the left arm and right arm (LA − RA). Lead II: potential difference between the left leg and right arm (LL − RA). Lead III: potential difference between the left leg and the left arm (LL − LA). LA: left arm; LL: left leg; RA: right arm.



XSL•FO
RenderX

Similar to the augmented leads, the absence of the WCT should theoretically prevent the recording of precordial leads. In this case, smartwatches create a closed positive/negative circuit at positions V1 to V6 [17], allowing the recording of each lead [5]. However, it should be noted that this configuration does not reproduce a true unipolar precordial lead referenced to the WCT, but rather generates a bipolar recording, which represents an approximation of conventional precordial leads [13].

Unlike conventional electrocardiographs, smartwatches cannot record the electrical activity of the heart simultaneously. With these devices, the leads are obtained sequentially, and to calculate the augmented leads, it is necessary to have previously recorded the limb leads (I, II, and III). By recording the leads sequentially, 3 independent electrocardiographic tracings are obtained, each with a slightly different heart rate.

Additionally, a representative case of Brugada syndrome recorded during a flecainide test was included as an illustrative example. This case was retrospectively selected from routine clinical practice and was not part of a predefined study protocol or a systematic pharmacological intervention.

ECG Signal Alignment and Augmented Lead Generation

Leads I, II, and III, recorded with the smartwatch, served as the basis for this analysis. From these, the augmented leads aVR, aVL, and aVF were calculated, considering only the frontal plane. The smartwatch provides sequential 30-second recordings, whereas conventional electrocardiographs generate 10-second segments. To ensure standardized comparison, the first 10 seconds of each lead obtained with the smartwatch were used. Subsequently, for the reconstruction of the 12-lead ECG, the signals were organized in a 3 × 4 format [18], following the conventional procedure in electrocardiography. In this format, 10 seconds of continuous recording are represented, divided into columns of 2.5 seconds: the first column starts at 0 seconds, the second at 2.5 seconds, the third at 5 seconds, and the fourth at 7.5 seconds. Thus, the electrical activity of all leads is displayed over a continuous 10-second interval, facilitating analysis and comparison with recordings obtained using standard electrocardiographs.

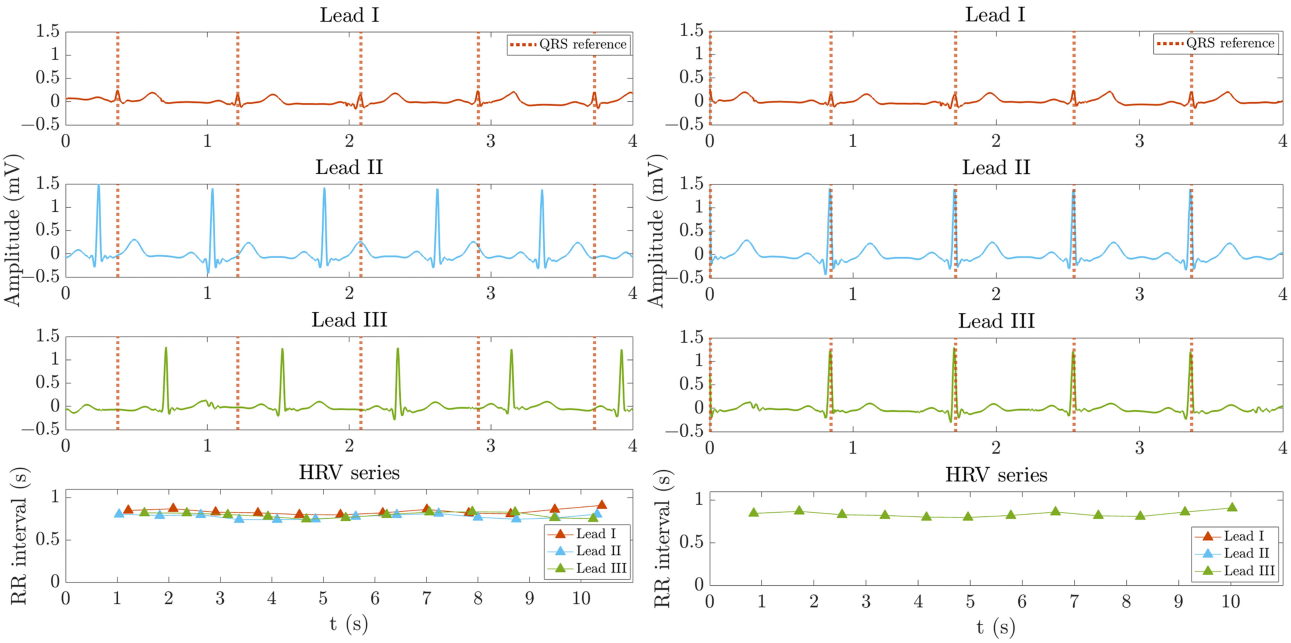
The first step involves detecting the QRS complexes and R peaks in each smartwatch lead using an AI-based detector developed by our group and previously validated [19]. All detections are manually reviewed in challenging cases to ensure accuracy.

In this study, lead I serves as the reference, with its R peak positions guiding the alignment of the other leads. We use the heart rate variability (HRV), which quantifies the temporal fluctuations between successive R peaks, to evaluate differences between leads. Thus, HRV is obtained by interpolating the interval function $d(t)$, which is considered to be a nonuniformly sampled signal of the variability function [20]:

$$(1)\quad d(t) = \sum_{k=1}^{B} (t_k - t_{k-1})\,\delta(t - t_k)$$

Here, t_k represents the time of the kth detected R peak, $\delta(t)$ is the Dirac delta function that indicates the location of each beat, and B is the number of beats in the sequence. Ideally, the HRV curves for all 3 leads would be the same, which is not the case due to sequential recording, as can be seen in the bottom-left graph of Figure 2.

Figure 2. ECG smartwatch leads I, II, and III (4 s out of the total 10 s window) and the corresponding HRV for the entire 10-second segment from a healthy control case. The left plot shows misaligned leads with nonoverlapping HRV curves. The right plot shows the aligned leads after the resampling process, with all HRV curves overlaid and aligned with each other. ECG: electrocardiogram; HRV: heart rate variability.



To align leads II and III with lead I, we resample the RR intervals to make them equal, ensuring that interbeat distances match those of lead I. This process is carried out by applying a sampling frequency change scheme to adjust the number of samples in each RR block and make them equal to the reference lead; it may be seen as an RR normalization procedure. Given that the resampling process is implemented for each RR block separately, we consider xl(s)[n], n=0, 1, ..., L_s− 1, as the sth RR

XSL•FO
RenderX

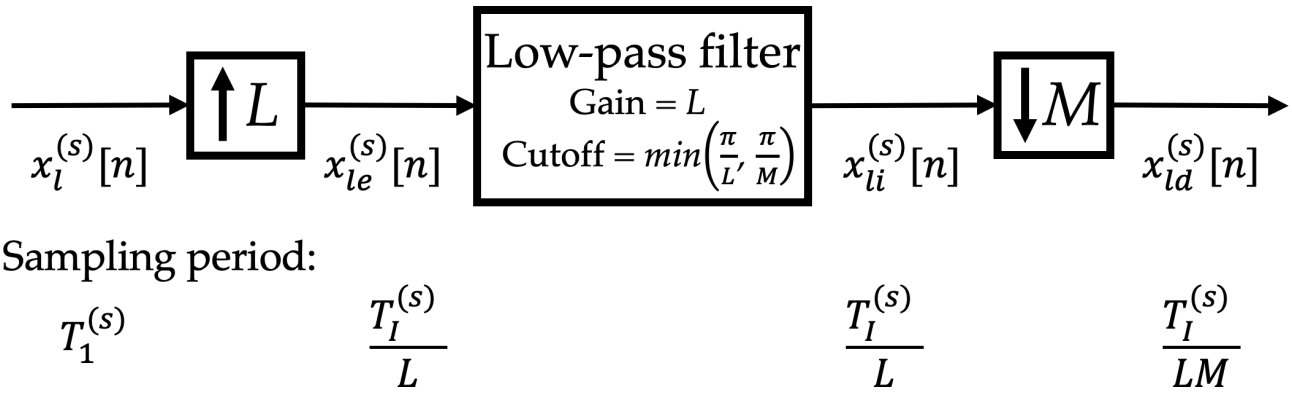
block of lead l, where $l \in$ {I, II, III}, so the ECG signal for lead l can be expressed as the concatenation of successive RR blocks:

(2)xl[n]= xl(0)[n]+ ∑s=1K−1xl(s)[n−∑i=0s−1Li]

where $K(B - 1)$ is the total number of RR intervals, and L_i is the length of the ith segment.

The resampling process, illustrated in Figure 3, involves three key steps [21]. First, each segment xl(s)[n] (for l=II, III) is upsampled by a factor L, increasing the number of samples. Next, a low-pass filter is applied to the upsampled segments to remove high-frequency artifacts introduced during interpolation, ensuring the signal remains within the Nyquist limit. Finally, the filtered segments are downsampled by a factor M to match the number of samples in lead I.

Figure 3. System for resampling by a noninteger factor, illustrating the combined processes of interpolation, filtering, and decimation.



This combination of interpolation, filtering, and decimation adjusts the sampling rate of each segment by a noninteger factor, ensuring that the RR intervals of leads II and III align with those of lead I. Importantly, the overall sampling frequency of the signal remains unchanged. By resampling each RR segment individually, this method preserves the temporal and morphological details of the ECG signal, ensuring that the QRS complexes and R peaks in leads II and III align with those in lead I. Since the ECG is oversampled, the waveform does not change visually. Thus, the amplitudes of the P, R, and T peaks remain the same, and the time scale is adjusted proportionally; that is, the scale is preserved across the entire interval. Unlike averaging techniques [11,12], which may mask pathological features, this approach minimizes information loss and maintains their integrity. Recent studies [13] have shown that smartwatch recordings obtained at various limb and precordial positions can yield amplitude and duration measurements comparable to standard 12-lead ECGs, but those signals are still acquired sequentially and compared lead-by-lead without alignment or RR-block resampling to synchronize morphology across leads. Similarly, Chen et al [14] propose a learning-based model that reconstructs a full 12-lead ECG from a single lead, generating new information to approximate the missing signals. In contrast, our method operates directly on the patient's recorded signals. RR-block resampling ensures interlead temporal consistency and preserves beat-level morphology, enabling the subsequent derivation of augmented leads. Once the primary leads are aligned, the augmented leads are calculated using Goldberger's equations [22]:

(3)aVR=−I+II2

(4)aVL=I−III2

(5)aVF=II+III2

This alignment ensures that each augmented lead is synchronized with the primary ECG signals, improving the reliability of subsequent analyses and enabling direct comparison with conventional 12-lead ECGs. The signal-processing script used to generate the augmented limb leads has been provided as Multimedia Appendix 1.

Obtaining 12-Lead ECG

The alignment process described for leads II and III can be generalized to include the precordial leads (V1-V6), enabling the reconstruction of a full 12-lead ECG. Using lead I as the reference, the same resampling methodology can be applied to align the precordial leads. This involves transforming the nonsynchronized RR intervals from leads V1 to V6 into synchronized representations, ensuring their cardiac intervals match those of lead I.

Let the sth block from leads V1-V6 be denoted by xvk(s)[n], where $v_k \in$ {V1, V2, V3, V4, V5, V6}. Each RR interval has a different period, Tvk(s), and the complete signals are defined as the concatenation of all segments (Equation 2). The same resampling process is applied to align the precordial leads with lead I. By aligning all leads (I, II, III, and V1-V6) obtained from the smartwatch, along with the generated augmented leads, the smartwatch ECG signals can be directly compared with conventional 12-lead ECG recordings. This generalization of the alignment process ensures that the precordial leads are synchronized with the primary leads, enabling a comprehensive analysis of the full 12-lead ECG. Such alignment improves the reliability of diagnostic comparisons between smartwatch-derived ECGs and standard clinical ECGs.

Clinical Validation

A prospective observational and comparative proof-of-concept study was conducted to validate augmented leads (aVR, aVL, and aVF) derived from Apple Watch Series 9 recordings against those from a conventional electrocardiograph. Twenty nonconsecutive cases were included. Of these, 19 patients admitted to cardiology presented diverse cardiac pathologies and evident pathological electrocardiographic alterations in the frontal plane. In addition, one healthy control in sinus rhythm

XSL•FO
RenderX

without known cardiac disease was included. The 12-lead ECGs were obtained using the previously described process.

It should be noted that the expert-based clinical validation in this study was limited to frontal plane leads, as these represent the primary focus of the proposed methodological contribution, namely the analytical reconstruction of augmented leads.

To complement this expert-based evaluation, a quantitative analysis was performed comparing conventional ECG (A) and smartwatch-derived ECG (B) across all cases. Standard electrocardiographic parameters were evaluated, including heart rate, PR interval, QRS duration, QT interval, QTc, RR interval, P-wave duration, and PP interval. All measurements were performed by experienced cardiology researchers using digital ECG analysis tools, ensuring consistent interval and waveform assessment across both modalities. For each parameter, mean values and SDs were calculated for both modalities across all cases, along with the mean difference between methods (A − B). This approach allowed the assessment of both central tendency and dispersion, providing an objective evaluation of agreement and variability between conventional and smartwatch-derived ECG recordings.

Additionally, a questionnaire was anonymously distributed to 85 health care professionals experienced in electrocardiographic diagnosis, comprising 50 cardiologists, 5 internists, 10 family physicians, 5 emergency physicians, and 15 critical care nurses. This comparative approach used data from both the Apple Watch Series 9 and a Philips TC30 conventional electrocardiograph (Koninklijke Philips N.V.) to assess the concordance and clinical utility of the smartwatch-generated augmented leads.

Questionnaire

To collect expert clinical opinions on the agreement and diagnostic concordance of electrocardiographic signals, a structured questionnaire was used. For each of the 20 cases, whose diagnoses were confirmed and supported by a clinical electrophysiologist, an image was presented with two ECG options, including only the frontal plane leads. These options, labeled “Option A” and “Option B,” randomly corresponded to either the conventional electrocardiograph recording or the smartwatch-generated tracing. The questionnaire comprised 5 key questions (Figure S1 in Multimedia Appendix 1) designed to assess specific aspects related to method agreement and experts’ perception of diagnostic concordance:

1. Diagnosis of electrocardiographic tracings: Evaluators selected from 20 available diagnostic options, acknowledging that some tracings might present multiple diagnoses due to coexisting alterations. This assessment included tracings from both smartwatches and conventional electrocardiographs.
2. Determination of the electrical axis of the cardiac: Evaluation and comparison of the electrical axis calculated using both technologies, based on the polarity of the QRS complex in leads I and aVF.
3. Diagnostic concordance between both technologies: Assessment of the level of agreement in diagnoses between smartwatch-derived ECGs and conventional ECG systems.
4. Visual correlation assessment of the tracings: Evaluators assigned a percentage score (1%-100%) to indicate the perceived degree of visual correlation between tracings obtained from both technologies. It should be noted that this “correlation score” represents a subjective visual assessment of morphological similarity and does not correspond to a mathematically computed correlation coefficient.
5. Identification of the tracing source (smartwatch or electrocardiograph): Assessment of clinicians’ ability to correctly distinguish between tracings obtained from a smartwatch and those from a traditional electrocardiograph.

The results were included in the centralized database for subsequent analysis.

Ethical Considerations

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki (World Medical Association, 2013) and its subsequent amendments. The study protocol was reviewed and approved by the Ethics Committee of Ramón y Cajal University Hospital (Madrid, Spain) on July 31, 2023 (approval number 181/23). All participants provided written informed consent prior to enrollment. Patient data were fully anonymized, and all study data were securely stored in compliance with applicable data protection regulations. No information that could reveal the identity of any participant has been included in this manuscript.

Results

Data supporting the findings of this study are provided in Multimedia Appendix 1.

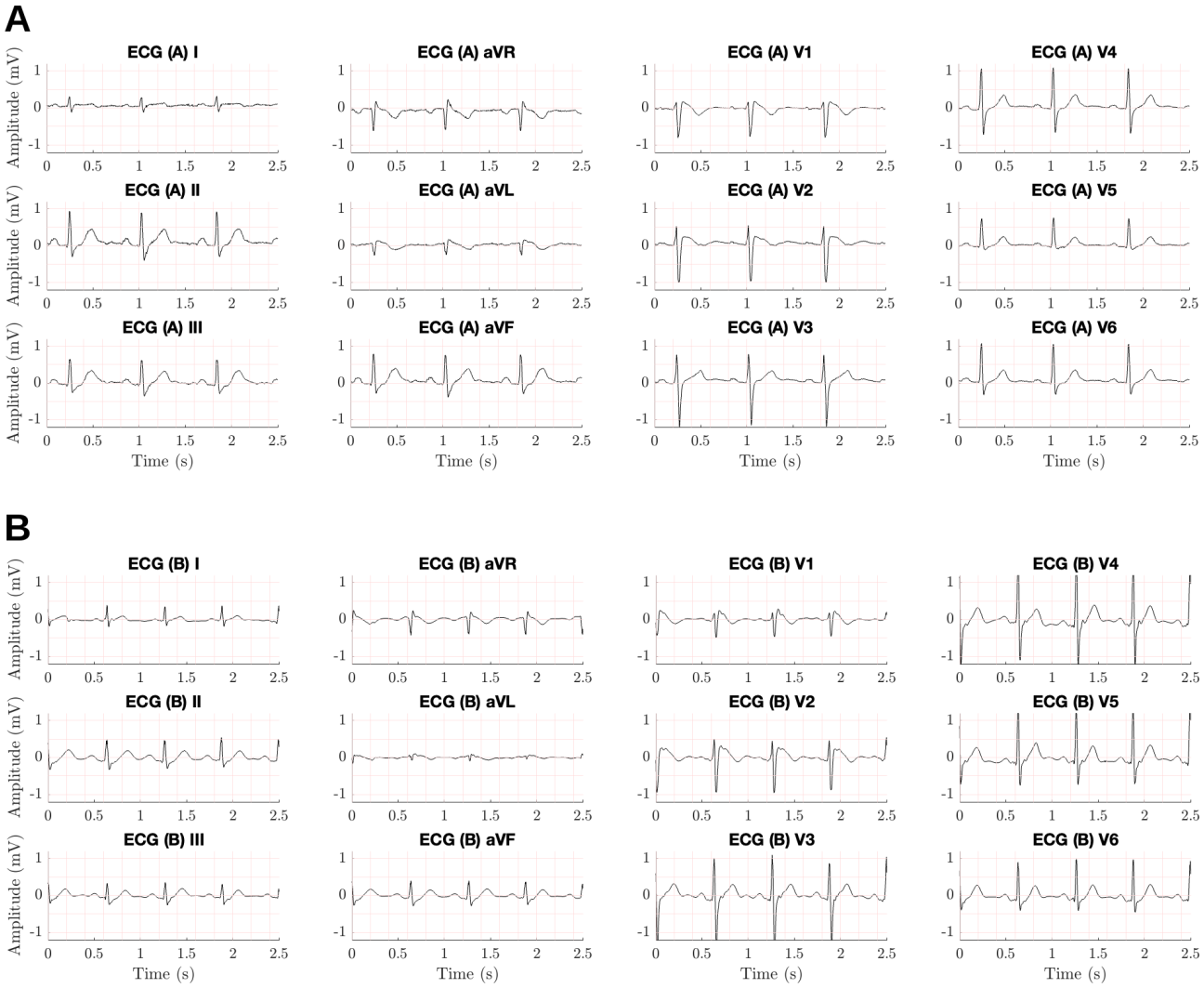
12-Lead ECG

To demonstrate an example of the methodology used to align the leads and generate augmented leads, we present a case of type I Brugada syndrome. Brugada syndrome is a cardiac pathology characterized by specific ECG patterns, including ST-segment elevation and T-wave inversions, and is associated with an increased risk of sudden cardiac death. This pathology provides an interesting scenario for comparing the 12-lead smartwatch-derived ECG with the conventional 12-lead ECG.

Figure 4 shows the results. Figure 4A shows the conventional 12-lead ECG recorded simultaneously from the same patient. Figure 4B displays the 12-lead ECG recorded with the smartwatch after alignment and generation of augmented leads, following the described methodology.

XSL•FO
RenderX

Figure 4. Brugada type I pattern during flecainide test. Conventional 12-lead ECG (A) and smartwatch 12-lead ECG (B). ECG: electrocardiogram.



Both ECGs are highly similar, demonstrating the accuracy of the reconstruction process. In both recordings, the characteristic Brugada type I pattern is clearly preserved, with typical ST-segment elevation in the right precordial leads (V1-V3). In addition, the corresponding morphology of the augmented limb leads (aVR, aVL, and aVF) is maintained, supporting the overall consistency of the reconstructed 12-lead ECG.

Clinical Validation

Each of the 20 cases underwent an independent review by an expert arrhythmologist. Detailed case-by-case clinical interpretations and representative figures are provided in Multimedia Appendix 1.

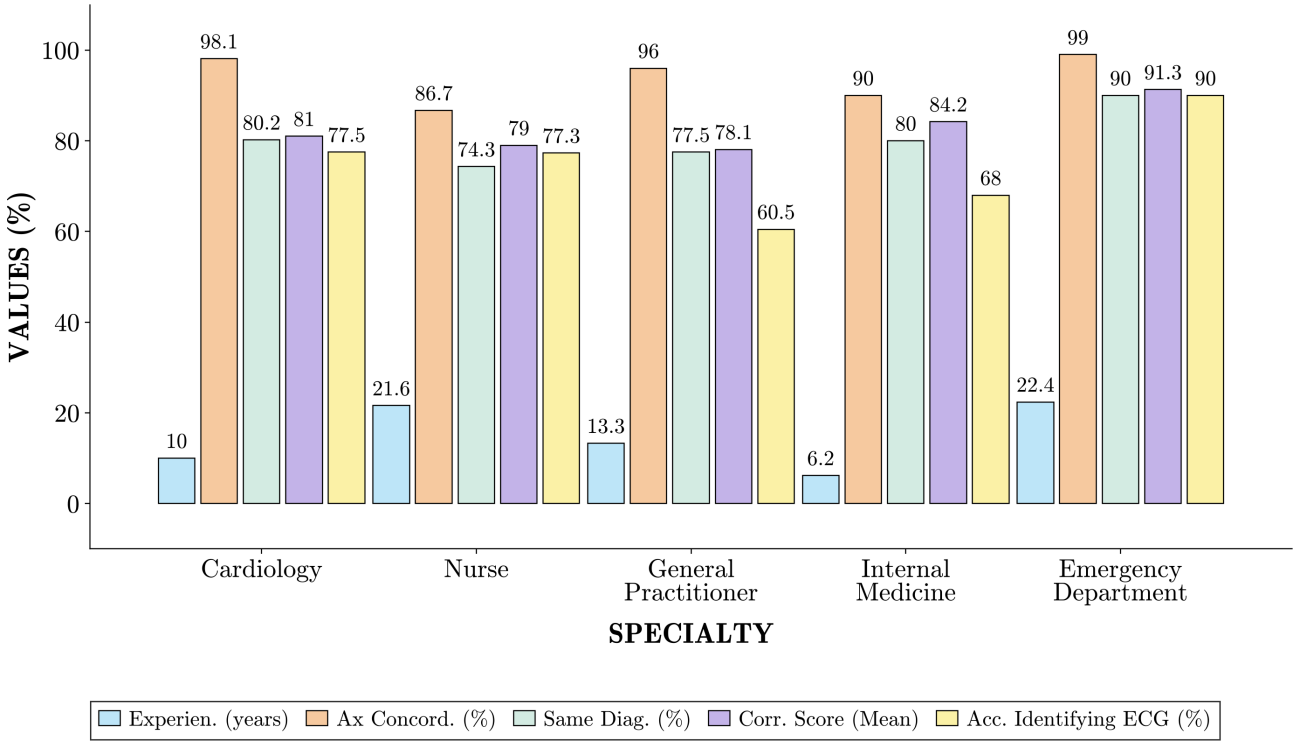
The analysis of data from the 20 cases studied shows variable outcomes depending on the specialties of the evaluators and the complexity of the cases. In general, diagnostic accuracy, axis concordance, and the consistency between diagnoses obtained from conventional leads and those recorded by the experimental device are consistent across specialties, although significant differences have been observed in certain cases.

The average perceived diagnostic concordance is 79.4%. Emergency specialists showed the highest agreement (90%), followed by cardiologists (80.2%) and internal medicine specialists (80.0%). General practitioners and nurses showed lower levels of agreement at 77.5% and 74.3%, respectively. Complex cases, such as cases 9 and 10, demonstrate lower general concordance, particularly in noncardiology specialties, reflecting the inherent difficulty in interpreting certain patterns.

The average evaluator-perceived concordance along the axis between diagnoses performed with the experimental device and the standard method is 95.4%, distinct from the investigators' direct axis measurements, which showed complete concordance in all cases. Emergency specialists show the highest concordance (99%), followed by cardiologists (98.1%) and general practitioners (96.0%). Although nurses and internal medicine specialists show slightly lower values (86.7% and 90%, respectively), they maintain high general concordance. The correlation scores range from 78.1 to 91.3 across specialty-level means, with an overall mean of 81.1 (SD 22.7). Emergency specialists show the highest scores (mean 91.3, SD 24.2), followed by internal medicine specialists (mean 84.2, SD 23.4) and cardiologists (mean 81.0, SD 22.2). Nurses and general practitioners obtain mean lower scores of 79.0 (SD 21.2) and 78.1 (SD 24.8), respectively (Figure 5). At the individual case level, scores ranged more widely, from 47.0 (case 10) to 96.3 (case 2).

XSL•FO
RenderX

Figure 5. Comparison of metrics by specialty. ECG: electrocardiogram.



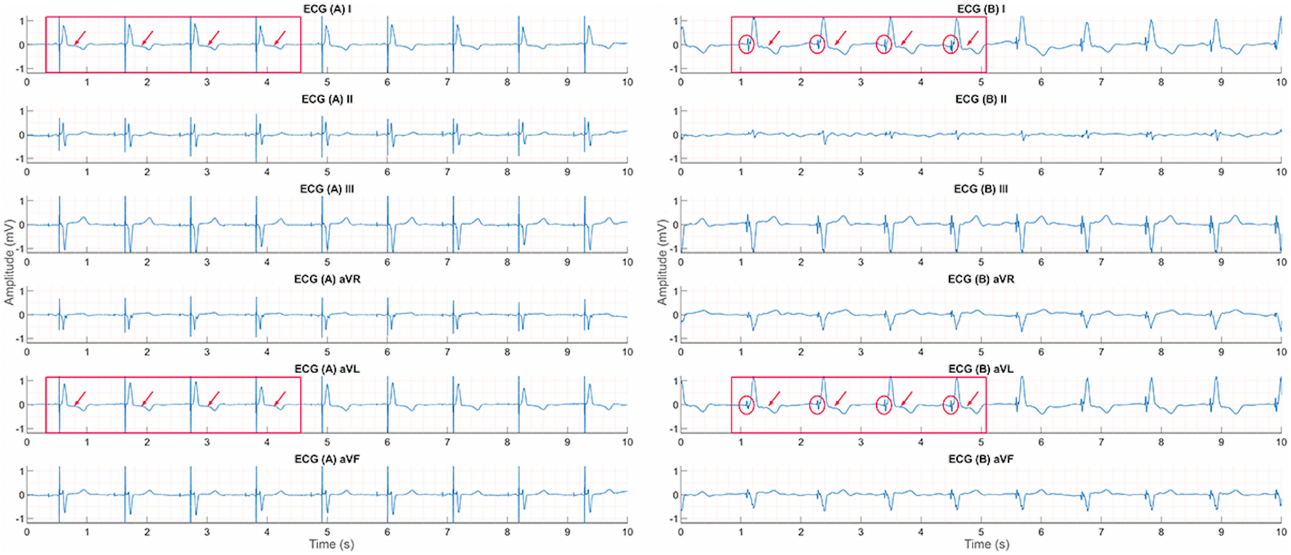
The correct identification of the source of the ECG (from a conventional source or from smartwatches) reaches an average of 75.6%. Emergency specialists show the highest accuracy (90%), followed by cardiologists (77.5%) and nurses (77.3%). General practitioners and internal medicine specialists show lower levels of accuracy (60.5% and 68%, respectively). As the similarity between the traces increases, the difficulty of correctly identifying the source of the ECG also increases. Among the 20 cases, one sinus rhythm ECG obtained from a healthy participant (case 2) is included as a control.

Before analyzing specific results, the cases were grouped by pathology to assess the behavior of the experimental device in different clinical scenarios. The thematic subgroups were not mutually exclusive; therefore, cases presenting more than one clinically relevant electrocardiographic abnormality could be included in more than one subgroup. Case 12, the only VT case in the cohort, was not included in the thematic subgroup analyses because no additional VT cases were available to constitute a homogeneous subgroup. It was included in the overall analysis and is presented individually in Multimedia Appendix 1 (Figure S13). In cases of ischemic heart disease (cases 6, 7, 13, 15, and 18), the concordance of the axis was high, with an average of 93.4%. Emergency specialists achieved 100%, while cardiologists reached 96.4%. Consistent diagnoses averaged 86.1%, with emergency specialists (96%) and cardiologists (87.2%) achieving the best results. The correlation scores had a mean of 84.9 (SD 19.2), with emergency physicians assigning the highest scores (95.2), followed by internal medicine specialists (90.0). The accuracy in identifying the source of the ECG was 76.0%, with emergency specialists standing out at 92% (Table S22 in Multimedia Appendix 1). In cases of pacemaker rhythms (cases 9, 10, and 14), the concordance of the axis was 94.5%; emergency specialists achieved 100% and cardiologists 98%. However, consistent diagnoses were lower, averaging 32.2%, and emergency specialists performed best at 46.7%. The correlation scores averaged 55.5 (SD 27.1), and internal medicine specialists achieved the highest scores (63.7). The accuracy in identifying the source of the ECG was high, averaging 93.7%, led by cardiologists (96.7%) and nurses (95.6%) (Table S23 in Multimedia Appendix 1).

The low correlation score between the tracings and the low percentage of identifying the same diagnosis were conditioned by the absence of pacemaker spikes in the smartwatch ECG tracing. In conventional electrocardiographs, a high-pass filter of 0.05 Hz and a low-pass filter of 150 Hz were configured, according to the recommendations of the ECG standardization and interpretation guidelines, published by the American Heart Association, the American College of Cardiology, and the Heart Rhythm Society [23]. The smartwatch, on the other hand, uses a more restrictive low-pass filter of around 30 to 45 Hz [24], which eliminates high-frequency signals such as pacemaker spikes without affecting the typical morphology of the pacemaker-stimulated ECG (Figure 6).

XSL•FO
RenderX

Figure 6. ECG A corresponds to the tracing obtained with a conventional electrocardiograph. In leads I and aVL, four cardiac cycles stand out, showing the inverted T-wave, a typical characteristic of the left bundle branch block pattern associated with pacemaker stimulation, with the ventricular electrode located in the right ventricle. On the other hand, ECG B is the tracing recorded by the smartwatch. In leads I and aVL, it is observed that it preserves the same morphology of left bundle branch block. The pacing spikes, which are circled to highlight them, appear attenuated due to the greater bandwidth restriction of the device. aVL: augmented voltage left; ECG: electrocardiogram.



To complement these findings, a quantitative comparison between conventional ECG (A) and smartwatch-derived ECG (B) was performed across all cases, and the results are summarized in Table 1 (details for each case are provided in Multimedia Appendix 1). Quantitative analysis demonstrated small mean differences between modalities, with particularly high agreement observed in structurally stable parameters such as heart rate (mean 77.2, SD 22.3 vs mean 75.9, SD 23.1 bpm; Δ=1.3 bpm) and PR interval (mean 176.2, SD 29.1 vs mean 180.8, SD 19.8 ms; Δ=−4.6 ms), computed only for cases in which each parameter was measurable. Greater variability was observed in time-dependent parameters such as QT (Δ=−19.2 ms) and QTc (Δ=−12.7 ms). For parameters representing atrial activation in eligible cases, the PP interval showed a mean of 859.0 (SD 175.5) ms vs 877.0 (SD 181.1) ms (Δ=−18.0 ms), while P-wave mean duration was 106.3 (SD 9.6) vs 119.6 (SD 13.9) ms (Δ=−13.3 ms). The electrical axis showed complete concordance between the two methods in all cases. Overall, structurally stable parameters such as heart rate and PR interval showed high agreement, whereas time-dependent parameters such as QT and QTc exhibited greater variability, with larger interval differences observed in some individual cases.

XSL•FO
RenderX

Table . Quantitative comparison of conventional electrocardiogram (ECG) (A) and smartwatch-derived ECG (B) across all cases in the cohort[a].

Parameter	ECG A, mean (SD)	ECG B, mean (SD)	Mean difference (A − B)
Heart rate (bpm)	77.2 (22.3)	75.9 (23.1)	1.3
PR[b] interval (ms)[c]	176.2 (29.1)	180.8 (19.8)	−4.6
QRS[d] duration (ms)	105.2 (19.4)	117.2 (24.2)	−12.0
QT interval (ms)	417.1 (55.7)	436.3 (55.4)	−19.2
QTc[e] (ms)	441.1 (32.5)	453.8 (33.2)	−12.7
RR[f] interval (ms)	843.7 (283.8)	878.0 (351.4)	−34.3
PP[g] interval (ms)[c]	859.0 (175.5)	877.0 (181.1)	−18.0
P-wave duration (ms)[c]	106.3 (9.6)	119.6 (13.9)	−13.3

[a]Data are presented as mean (SD). Mean differences were calculated as ECG A minus ECG B. The electrical axis showed complete concordance between the two methods in all cases.

[b]PR: P-R interval.

[c]Calculated exclusively from cases with physiologically present and measurable waveforms, to avoid zero-value calculation bias in nonapplicable conditions. Number of contributing cases (ECG A/ECG B): 12/11 (PR interval), 12/11 (PP interval), and 15/14 (P-wave duration). Where a parameter was measurable in one modality but not in the other, the case contributes only to the modality in which it could be determined; the mean difference therefore represents the difference between the modality-specific means. Exclusions were applied as follows: (1) PR interval (ECG A, n=12; ECG B, n=11) was excluded owing to absent or unmeasurable atrioventricular conduction in complete AV block (case 1), second-degree AV block 2:1 (case 4), atrial fibrillation/flutter (cases 5, 17, 19, and 20), ventricular tachycardia (case 12), and second-degree AV block Mobitz I (case 18); additionally, it was excluded from ECG B only in the ventricularly paced rhythm with nondiscernible P waves (case 9, ECG B only). (2) PP interval (ECG A, n=12; ECG B, n=11) was excluded owing to absent, independent, or highly irregular/fibrillatory atrial rhythms in complete AV block (case 1), second-degree AV block 2:1 (case 4), atrial fibrillation/flutter (cases 5, 17, 19, and 20), ventricular tachycardia (case 12), and second-degree AV block Mobitz I (case 18); additionally, it was excluded from ECG B only in the ventricularly paced rhythm (case 9, ECG B only). (3) P-wave duration (ECG A, n=15; ECG B, n=14) was excluded owing to absent or nondiscernible P waves in atrial fibrillation/flutter (cases 5, 17, 19, and 20) and ventricular tachycardia (case 12); additionally, it was excluded from ECG B only in the ventricularly paced rhythm with nondiscernible P waves (case 9, ECG B only).

[d]QRS: QRS complex.

[e]QTc: corrected QT interval.

[f]RR: R-R interval.

[g]PP: P-P interval.

In cases of atrioventricular or bundle branch blocks (cases 1, 3, 4, 6, 8, 11, 16, and 18), the concordance of the axis averaged 93.5%, with emergency specialists at 97.5% and cardiologists at 96.8%. Consistent diagnoses reached 84.4%, with emergency specialists standing out at 97.5% and cardiologists at 87.2%. The correlation score averaged 82.6 (SD 20.0), and emergency specialists granted the highest scores (96.8%). The accuracy in identifying the source of the ECG was 71.2%, with emergency specialists placed at 85.0% (Table S24 in Multimedia Appendix 1).

In cases of atrial arrhythmias (cases 5, 17, 19, and 20), the concordance of the axis was averaged at 96.8%, with emergency specialists achieving 100% and cardiologists 99.5%. Consistent diagnoses averaged 86.8%, with emergency specialists (100%) and internal medicine specialists (95.0%) showing the highest agreement. The correlation scores averaged 86.1 (SD 17.3), and emergency specialists granted the highest score (99.8%). Accuracy in identifying the source of the ECG was 69.7%, with emergency specialists achieving 90% (Table S25 in Multimedia Appendix 1).

Ischemic heart disease, atrial arrhythmias, and atrioventricular or bundle branch blocks demonstrated the highest diagnostic concordance. Conversely, pacemaker rhythms exhibited challenges for intermethod correlation and consistency (Figure 6). Emergency physicians consistently achieved superior diagnostic accuracy.

Discussion

Principal Findings

This study validated the reconstruction of 12-lead ECGs from smartwatch-recorded signals by aligning leads and generating augmented leads. Our findings demonstrate high concordance between smartwatch-derived ECGs and conventional ECGs, particularly in ischemic heart disease and bundle branch blocks, supporting their agreement with conventional ECGs and their potential as a complementary tool for cardiac condition evaluation.

In addition, quantitative analysis of electrocardiographic parameters demonstrated small mean differences between modalities, with particularly high agreement in structurally stable measurements such as heart rate and PR interval, which supports the temporal consistency of the reconstructed signals despite sequential acquisition.

The alignment of leads and the generation of augmented leads proved feasible and effective, yielding clinically meaningful results. It is important to emphasize that the methodological innovation of this study primarily concerns the reconstruction

XSL•FO
RenderX

of frontal plane leads, while precordial leads are directly acquired and subsequently synchronized; therefore, clinical validation was focused on the frontal plane, where the core methodological contribution lies. This innovative approach, using sequential tracings and advanced signal processing, overcomes previously identified technological barriers by enabling reconstruction of augmented leads with clinically meaningful electrocardiographic information.

Comparison With Prior Work

Previous studies have explored alternative strategies for extending smartwatch ECG capabilities, including beat averaging to generate representative ECG complexes [11,12], direct sequential comparison of smartwatch-acquired leads without interlead synchronization [13], and learning-based generative reconstruction methods designed to synthesize missing ECG leads from partial input signals [14].

In contrast, our approach preserves beat-level morphology by aligning RR intervals across sequential recordings and analytically deriving the augmented leads, rather than generating synthetic electrocardiographic information. This distinction is particularly relevant because smartwatch recordings are inherently asynchronous, creating a specific methodological challenge that differs substantially from conventional ECG reconstruction settings.

In the case of learning-based generative models [14], the objective is typically to reconstruct a full 12-lead ECG from one or more conventional ECG leads by solving a data-driven synthesis problem. Since these reconstructed signals are synchronized with the original acquisition, waveform similarity metrics can be used to assess reconstruction fidelity. However, the applicability of this methodology to wearable devices remains uncertain, as real-world smartwatch acquisition introduces additional constraints that are not typically represented in existing experimental frameworks.

An important distinction is that lead I, as defined in conventional electrocardiography, is not directly available on wearable devices due to physical hardware constraints, as the electrodes are positioned in very close proximity to each other. Therefore, the starting signal obtained from a smartwatch represents only an approximation of conventional lead I rather than a true equivalent.

Our methodology differs conceptually in that it operates directly on actual patient-recorded smartwatch signals, using RR-block resampling to address the challenge of sequential, asynchronous acquisition while preserving temporal consistency across leads. Rather than reconstructing an approximation from synthetic data, the proposed framework builds the 12-lead ECG from real electrocardiographic measurements acquired under wearable conditions.

Clinical Implications

The increasing commercial availability and adoption of smartwatches have positioned these devices as potentially valuable tools in cardiovascular medicine. While their current clinical utility is primarily established in atrial fibrillation detection through single-lead ECG acquisition [25], the inherent limitation of a single lead restricts their applicability in broader diagnostic scenarios, such as ischemic heart disease, myocardial pathology, or accurate arrhythmia localization.

The methodology presented in this study expands this diagnostic scope by enabling reconstruction of a 12-lead ECG from smartwatch-acquired signals. This aligns with growing interest in smartwatch-based cardiac assessment for ischemic heart disease [7], including our group's prior work demonstrating the feasibility of multilead smartwatch ECG acquisition [6,26]. Our group is also currently conducting a large-scale multicenter study to evaluate an abbreviated 3-lead smartwatch algorithm for ischemic heart disease detection [27].

Our analysis further showed that emergency physicians and cardiologists ranked first and second, respectively, in accuracy in identifying the ECG source and in axis concordance. For the correlation score, however, emergency physicians ranked highest, followed by internal medicine specialists, with cardiologists in third place. This finding highlights the importance of clinical expertise in interpreting complex electrocardiographic tracings and suggests that smartwatch-derived ECGs may have practical value as diagnostic adjuncts when integrated into established clinical workflows.

Future Directions

It should be noted that the reconstructed ECGs in this study were not presented in a standard clinical display format, as the primary objective was to demonstrate the feasibility of the proposed methodology as a proof-of-concept. The visualization layout used represents an initial approximation for comparative purposes; however, the underlying signal processing framework is flexible and can be adapted to conventional ECG display formats, including those incorporating a continuous rhythm strip. Future developments will focus on implementing standardized clinical layouts to enhance interpretability and practical applicability.

Beyond technical implementation, further research is needed to evaluate the clinical applicability of smartwatch-derived 12-lead ECG reconstruction in larger and pathology-specific cohorts. Smartwatch-based ECG technologies may also prove useful in broader remote cardiac monitoring scenarios, including rhythm surveillance and bradyarrhythmia detection, as suggested by previously published case reports [28]. Broader validation will be essential to determine performance across different cardiovascular conditions and real-world monitoring settings.

Future integration with AI may further enhance wearable ECG analysis through individualized longitudinal monitoring, automated anomaly detection, and decision-support tools [29]. These applications remain exploratory and require rigorous clinical validation before adoption in routine practice.

Limitations

Despite the promising results, several limitations warrant consideration, particularly given the proof-of-concept nature of this study. First, the current methodology involves a manual and potentially subjective visual review for optimal segment selection from smartwatch recordings. However, this is a known

XSL•FO
RenderX

challenge with clear pathways for automation through signal quality indices in future iterations. Similarly, while a learning-based R peak detector was used [19], complex pathologies, including pacemaker rhythms, occasionally required manual review due to inherent challenges in interpretation and consistency. This reliance on manual intervention, although appropriate for early-stage validation, may limit scalability and real-world applicability, highlighting the need for fully automated signal processing workflows in future implementations. These issues, common in early-stage development, underscore the need for optimized recording techniques and improved filtering in portable devices, rather than representing insurmountable barriers.

Second, the sequential nature of smartwatch ECG acquisition inherently limits the applicability of conventional signal fidelity metrics, such as Pearson correlation coefficients, root mean square error, or signal-to-noise ratio. Unlike standard ECG systems, where all leads are recorded simultaneously, smartwatch-derived leads are obtained from different cardiac cycles and subsequently aligned through resampling. Therefore, direct point-by-point waveform comparison would not reflect true reconstruction accuracy but rather physiological beat-to-beat variability and temporal misalignment. For this reason, the present study focused on clinically meaningful interval-based comparisons and expert interpretation, which better reflect real-world diagnostic applicability.

Third, sequential acquisition may result in recordings being obtained during different phases of the respiratory cycle, which can alter the anatomical position of the heart and lead to variations in the electrical axis and QRS amplitude. Consequently, the application of the Goldberger equations to signals acquired under different respiratory conditions may introduce discrepancies compared to true simultaneous recordings.

Fourth, the limited frequency bandwidth of current smartwatch ECG devices represents an additional technical constraint, particularly for the detection of high-frequency components such as pacemaker spikes. Commercially available smartwatches typically operate within a reduced bandwidth (approximately 0.5 - 40 Hz), compared to conventional electrocardiographs (0.05 - 150 Hz), due to the application of relatively aggressive filtering, which may attenuate high-frequency signals and contribute to reduced diagnostic consistency in specific scenarios. This limitation specifically affects the identification of pacing artifacts, although the overall morphology of paced rhythms is generally preserved.

Fifth, precordial leads obtained with the smartwatch should be interpreted as approximations rather than as true unipolar recordings due to the absence of a WCT, which represents an inherent electrophysiological limitation of the device.

Sixth, the validation was conducted on a cohort of 20 cases, including patients with different cardiac conditions and one healthy control. Although the cohort included a range of electrocardiographic findings, some conditions were represented by more than one case (eg, atrial fibrillation and atrial flutter). While this limits generalizability, it is appropriate for a proof-of-concept study focused on validating the proposed methodology for 12-lead ECG reconstruction. In this context, the primary aim of the study was to demonstrate the technical feasibility and preliminary clinical agreement of the proposed reconstruction methodology under controlled conditions, rather than to establish diagnostic performance across specific pathologies. Therefore, larger studies including a greater number of patients within each diagnostic category will be necessary to further evaluate the clinical applicability and generalizability of the proposed methodology.

Seventh, the study cohort was intentionally composed of patients with evident electrocardiographic abnormalities, which may introduce a degree of selection bias and limit the extrapolation of results to real-world populations with varying pretest probabilities.

Therefore, the findings should not be interpreted as reflecting predictive or screening performance in unselected real-world populations, particularly in settings with lower pretest probability or more subtle electrocardiographic alterations.

Eighth, although the validation relied on expert opinion, a clinical electrophysiologist served as the “gold standard” to guide ECG interpretation, thereby acknowledging the inherent subjective component of ECG reading. To further strengthen these findings, more standardized and objective validation methods are needed in subsequent research.

Ninth, the use of anonymized survey responses may introduce response bias, and the absence of variability observed in certain subgroups should be interpreted with caution, as individual responses could not be verified. Furthermore, differences in subgroup sizes across specialties may influence the aggregated results, and although weighted analyses were used, residual imbalances may persist.

Tenth, to fully establish clinical utility, larger-scale studies focusing on specific pathologies are essential. Each cardiac condition presents unique diagnostic considerations, and the impact of early diagnosis varies significantly, from life-saving interventions in some cases to less critical implications in others. Therefore, future research should meticulously investigate the performance of smartwatch-derived ECGs across diverse cardiac conditions to ensure their reliability and appropriate clinical application.

Conclusions

This proof-of-concept study demonstrates the technical feasibility of reconstructing a synchronized 12-lead ECG from smartwatch-acquired signals through sequential lead alignment and analytical generation of augmented leads.

The proposed methodology showed preliminary clinical agreement with conventional ECGs, particularly in cases of ischemic heart disease and bundle branch blocks, while preserving beat-level morphology across sequential recordings.

These findings support the potential applicability of smartwatch-derived multilead ECG reconstruction as a complementary approach for wearable cardiac monitoring under controlled acquisition conditions.

XSL•FO
RenderX

Acknowledgments

The authors are grateful to the following for their support of this project through data collection: Lorena Piña Astete, Laura Freire Pernía, Jesús García Montalvo, Lidia Carreto García, María Irene Muñoz Gómez, Nerea Peñaranda Romero, Francisco Ruperto Gámez, Ana Tur Sainz, Luis Manuel Domínguez Rodríguez, Cardiology Department, Hospital Universitario Ramón y Cajal, Madrid, Spain. ChatGPT by OpenAI was used for grammar correction during manuscript generation.

Funding

This work has been partially supported by research project grants PID2022-140786NB-C32 by MCIN/AEI/10.13039/501100011033 and PRE FPU-UAH-21 from the University of Alcalá. The publication of this article was supported by the Ramón y Cajal Institute for Health Research (IRYCIS) and the University of Alcalá.

Data Availability

The data supporting the findings of this study are stored in an institutional database and are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MBS, GLAS, CP-S, MB-V
Data curation: MBS, GLAS, CP-S, MB-V
Formal analysis: MBS, CP-S
Funding acquisition: MB-V
Investigation: MBS, GLAS, CP-S, MB-V, RMF
Methodology: MBS
Resources: GLAS, CP-S, MB-V
Software: CP-S
Supervision: GLAS
Validation: MBS, GLAS, CP-S, MB-V, RMF
Visualization: MBS
Writing – original draft: MBS, CP-S, MB-V
Writing – review & editing: GLAS, CP-S, MB-V

Conflicts of Interest

The authors declare a pending European patent (EP25382712.5) related to the methodology described in this work. This research was conducted within public academic and health care institutions, including Ramón y Cajal University Hospital, the Ramón y Cajal Institute for Health Research (IRYCIS), the University of Alcalá, and the University Hospital of Navarra. No financial compensation has been received, and the patent did not influence the study design, data collection, analysis, or interpretation of the results.

Multimedia Appendix 1
Detailed clinical case analyses, representative ECG figures, quantitative electrocardiographic measurements, the signal-processing script used for augmented limb lead generation, and survey results.
[PDF File, 1837 KB - cardio_v10i1e86621_app1.pdf]

References

1. Samol A, Bischof K, Luani B, Pascut D, Wiemer M, Kaese S. Single-lead ECG recordings including Einthoven and Wilson leads by a smartwatch: a new era of patient directed early ECG differential diagnosis of cardiac diseases? Sensors (Basel) 2019 Oct 10;19(20):10. [doi: 10.3390/s19204377] [Medline: 31658713]
2. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
3. Lubitz SA, Faranesh AZ, Selvaggi C, et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit Heart Study. Circulation 2022 Nov 8;146(19):1415-1424. [doi: 10.1161/CIRCULATIONAHA.122.060291] [Medline: 36148649]
4. Campo D, Elie V, de Gallard T, et al. Atrial fibrillation detection with an analog smartwatch: prospective clinical study and algorithm validation. JMIR Form Res 2022 Nov 4;6(11):e37280. [doi: 10.2196/37280] [Medline: 35481559]

XSL•FO
RenderX

5. Barrera N, Solorzano M, Jimenez Y, Kushnir Y, Gallegos-Koyner F, Dagostin de Carvalho G. Accuracy of smartwatches in the detection of atrial fibrillation: a systematic review and diagnostic meta-analysis. JACC Adv 2025 Nov;4(11 Pt 1):102133. [doi: 10.1016/j.jacadv.2025.102133] [Medline: 41182224]
6. Buelga Suárez M, Pascual Izco M, Pastor Pueyo P, Lozano Granero C, García Montalvo J, Alonso Salinas GL. Smartwatch ECG tracing and ischemic heart disease: ACS Watch Study. Cardiology 2023;148(1):78-82. [doi: 10.1159/000528116] [Medline: 36412561]
7. Spaccarotella CAM, Polimeni A, Migliarino S, et al. Multichannel electrocardiograms obtained by a smartwatch for the diagnosis of ST-segment changes. JAMA Cardiol 2020 Oct 1;5(10):1176-1180. [doi: 10.1001/jamacardio.2020.3994] [Medline: 32865545]
8. Guarnieri G, Mapelli M, Moltrasio M, Agostoni P, Tondo C. Connected health: ventricular tachycardia detection with Apple Watch—a case report. Heliyon 2024 Dec 15;10(23):e40595. [doi: 10.1016/j.heliyon.2024.e40595] [Medline: 39654722]
9. Burke J, Haigney MCP, Borne R, Krantz MJ. Smartwatch detection of ventricular tachycardia: case series. HeartRhythm Case Rep 2020 Oct;6(10):800-804. [doi: 10.1016/j.hrcr.2020.08.003] [Medline: 33101960]
10. Madias JE. A proposal for a 9-lead electrocardiogram recorded via the Wilson's central terminal. Ann Noninvasive Electrocardiol 2013 Mar;18(2):103-106. [doi: 10.1111/anec.12040] [Medline: 23530479]
11. Ugander M, Curtis T. A method and a system for producing a standard 12-lead ECG. 2023 Nov 16 URL: https://patents.google.com/patent/US20230363684A1/en?oq=US20230363684A1 [accessed 2026-10-06]
12. Albert DE. Two electrode apparatus and methods for twelve lead ECG. 2016 May 31 URL: https://patents.google.com/patent/US9351654B2/en?oq=US9351654B2 [accessed 2026-10-06]
13. van der Zande J, Strik M, Dubois R, et al. Using a smartwatch to record precordial electrocardiograms: a validation study. Sensors (Basel) 2023 Feb 25;23(5):2555. [doi: 10.3390/s23052555] [Medline: 36904759]
14. Chen J, Wu W, Liu T, Hong S. Multi-channel masked autoencoder and comprehensive evaluations for reconstructing 12-lead ECG from arbitrary single-lead ECG. NPJ Cardiovasc Health 2024 Dec 4;1(1):34. [doi: 10.1038/s44325-024-00036-4] [Medline: 41775924]
15. Barold SS. Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. Card Electrophysiol Rev 2003 Jan;7(1):99-104. [doi: 10.1023/a:1023667812925] [Medline: 12766530]
16. Avila CO. Novel use of Apple Watch 4 to obtain 3-lead electrocardiogram and detect cardiac ischemia. Perm J 2019;23:19-025. [doi: 10.7812/TPP/19-025] [Medline: 31314734]
17. Gamble P, McManus H, Jensen D, Froelicher V. A comparison of the standard 12-lead electrocardiogram to exercise electrode placements. Chest 1984 May;85(5):616-622. [doi: 10.1378/chest.85.5.616] [Medline: 6713970]
18. PageWriter TC20/TC30 electrocardiographs: instructions for use [Report in Spanish]. : Philips Healthcare; 2011 URL: https://blogs.hsjda.es/electromedicina/wp-content/uploads/sites/10/2019/04/03-01-094-677-0X-C00001-USO-ECG-Spanish_TC20_TC30_IFU.pdf [accessed 2026-10-06]
19. Plaza-Seco C, Barner KE, Holgado-Cuadrado R, Melgarejo-Meseguer FM, Rojo-Álvarez JL, Blanco-Velasco M. Detection of intra-beat waves on ambulatory ECG using manifolds: an explainable deep learning approach. Biomed Signal Process Control 2025 Jul;105:107523. [doi: 10.1016/j.bspc.2025.107523]
20. Sörnmo L, Laguna P. Bioelectrical Signal Processing in Cardiac and Neurological Applications: Academic Press; 2005.
21. Oppenheim AV, Schafer RW. Discrete-Time Signal Processing, 3rd edition: Pearson; 2021.
22. Madias JE. On recording the unipolar ECG limb leads via the Wilson's vs the Goldberger's terminals: aVR, aVL, and aVF revisited. Indian Pacing Electrophysiol J 2008;8(4):292-297. [Medline: 18982138]
23. Kligfield P, Gettes LS, Bailey JJ, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. Circulation 2007 Mar 13;115(10):1306-1324. [doi: 10.1161/CIRCULATIONAHA.106.180200] [Medline: 17322457]
24. Knecht S, Waldmann F, Kuhn R, et al. Technical characterization of single-lead ECG signals from 4 different smartwatches and its potential clinical implications. JACC Clin Electrophysiol 2023 Aug;9(8 Pt 1):1415-1417. [doi: 10.1016/j.jacep.2023.03.004] [Medline: 37074248]
25. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 1;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
26. Buelga Suárez M, García Montalvo J, Alonso Salinas GL. Smartwatches and ischemic heart disease. A future possibility? Rev Esp Cardiol (Engl Ed) 2022 Sep;75(9):764. [doi: 10.1016/j.rec.2022.02.005] [Medline: 35305946]
27. Buelga Suárez M, Pascual Izco M, García Montalvo J, Alonso Salinas GL. Accuracy of smartwatch electrocardiographic recording in the acute coronary syndrome setting: rationale and design of the ACS WATCH II Study. J Clin Med 2024 Jan 10;13(2):389. [doi: 10.3390/jcm13020389] [Medline: 38256523]

XSL•FO
RenderX

28. Buelga Suárez ML, Ramírez Martín J, Alonso Salinas GL. Smartwatch health monitoring in the context of a complete heart block: a challenge beyond these watches' capabilities. Emergencias 2023 Dec;35(6):478-479. [doi: 10.55633/s3me/E019.2023] [Medline: 38116979]
29. Lim J, Park G, Lee HS, et al. ECG trained artificial intelligence for the detection of patients with inducible myocardial ischemia. Eur Heart J Digit Health 2026 Apr;7(3):ztag050. [doi: 10.1093/ehjdh/ztag050] [Medline: 41929849]

Abbreviations

AF: atrial fibrillation
aVF: augmented voltage foot
aVL: augmented voltage left
aVR: augmented voltage right
ECG: electrocardiogram
HRV: heart rate variability
QTc: corrected QT interval
VT: ventricular tachycardia
WCT: Wilson Central Terminal

Edited by A Coristine; submitted 27.Oct.2025; peer-reviewed by G Yang, Y Kubota; revised version received 05.Sep.2026; accepted 09.Sep.2026; published 05.Nov.2026.

Please cite as:
Buelga Suárez M, Plaza-Seco C, Matía Francés R, Alonso Salinas GL, Blanco-Velasco M
Generation of 12-Lead Electrocardiogram Using Smartwatches in Patients With Cardiac Conditions: Clinical Validation Study
JMIR Cardio 2026;10:e86621
URL: https://cardio.jmir.org/2026/1/e86621
doi:10.2196/86621

© Mauro Buelga Suárez, Carmen Plaza-Seco, Roberto Matía Francés, Gonzalo Luis Alonso Salinas, Manuel Blanco-Velasco. Originally published in JMIR Cardio (https://cardio.jmir.org), 5.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps for Heart Rate Variability: App Store Search and Content Analysis

Eline de Jager[1], MSc; Brian Caulfield[1,2], PhD; Evgenia Angelidi[1,2], MSc; Sinead Holden[1,3], PhD

[1]School of Public Health, Physiotherapy, and Sports Science, University College Dublin, Belfield, Dublin, Ireland

[2]Insight Research Ireland Centre for Data Analytics, Dublin, Ireland

[3]UCD Institute for Sport and Health, University College Dublin, Dublin, Ireland

Corresponding Author:
Eline de Jager, MSc
School of Public Health, Physiotherapy, and Sports Science, University College Dublin, Belfield, Dublin, Ireland

Abstract

Background: Heart rate variability (HRV) is a noninvasive indicator of autonomic nervous system activity that is increasingly used for health and performance monitoring. Digital and mobile technologies are increasingly providing opportunities for remote HRV monitoring outside of laboratory-based settings.

Objective: This study aimed to describe the landscape of mobile apps that measure, analyze, and provide feedback on HRV, with a focus on how HRV is measured, analyzed, interpreted, and communicated to users. A secondary aim was to assess the transparency of these apps, including the extent to which they disclose the evidence underpinning their HRV metrics and feedback.

Methods: This study was an app store search and content analysis. Searches were conducted in the Google Play Store and Apple iTunes Store. Apps were eligible for inclusion if they had functionality to record, analyze, or provide feedback on HRV and were available in English. Data were extracted from app descriptions, screenshots, websites, and, where necessary, contact with developers. Data were extracted on app metadata (developer, release and update dates, and pricing), alongside information about HRV measurement, analysis, and feedback. This included the type of sensor used; HRV measurement characteristics (sensor placement, recording duration, and body position or standardization procedures); methods to calculate and interpret HRV (ie, metrics derived and how they were interpreted for users); and additional app functionality such as reminders, the ability to log self-reported stressors, and the type of feedback or guidance provided based on HRV. We used previously published criteria for assessing the quality of information on the internet, which included authorship, scientific attribution, currency of updates, and data privacy.

Results: Of 746 apps identified, 206 met eligibility criteria. Of these, 132 were primary measurement apps, 59 were aggregators, and 15 were hybrid. Photoplethysmogram was the most common sensing modality (n=117, 56.8%), followed by multiple sensors (n=60, 29.1%). Full data extraction across app metadata and HRV measurement and analysis data was only achievable for 93 (45.1%) apps, representing a transparent subset with sufficient available information for content analysis. The most commonly reported HRV metrics were root mean square of successive differences (n=51) and SD of normal-to-normal intervals (n=48), while frequency-domain power (n=22) and low frequency to high frequency ratios (n=15) were less common. Most apps presented data as personalized trends or individualized ranges (76/93, 81.7%), emphasizing user-specific context rather than isolated values. Although 86% (80/93) offered contextual guidance (eg, readiness or recovery scores), many relied on proprietary algorithms that were not transparently described, limiting independent assessment of how these scores were derived and validated.

Conclusions: Consumer HRV apps are widely available but vary considerably in how data are collected, processed, and contextualized. While many offer personalized trends and guidance, methodological transparency is often limited, particularly regarding the proprietary algorithms underlying the feedback scores.

(JMIR Cardio 2026;10:e84764) doi:10.2196/84764

KEYWORDS

smartphone apps; photoplethysmography; digital health; wearable electronic devices; mobile health; mHealth; autonomic nervous system; devices

Introduction

Heart rate variability (HRV) is the variation in time intervals between consecutive heartbeats [1,2]. This metric is a noninvasive indicator of autonomic nervous system activity, reflecting the balance between the sympathetic (fight-or-flight) and parasympathetic (rest-and-digest) branches [1,2]. Generally, higher HRV is associated with greater parasympathetic influence, indicating effective recovery, adaptability, and overall

XSL•FO
RenderX

cardiovascular health [3-5]. Conversely, lower HRV is often linked to increased sympathetic activity, which may result from physical stress, intense training, illness, or other physiological and psychological factors [3-5]. HRV metrics have been shown to correlate with a range of health outcomes, including cardiovascular fitness, stress levels, and overall well-being [3-5].

The gold standard for measurement of HRV is an electrocardiogram (ECG), with traditional assessment protocols ranging from 5 minutes to 24 hours [1]. Many smartphone apps now use photoplethysmography (PPG) technology to estimate HRV by detecting changes in blood volume associated with each heartbeat through light reflected from the skin and captured by smartphone cameras or wearable sensors [6]. This enables an estimate of HRV through the pulse rate waveform. This has made HRV monitoring highly accessible to the general public. Advances in signal processing have enabled PPG to achieve reasonable accuracy under resting conditions [7-9]. However, this accessibility has created a "transparency gap": while many apps provide sophisticated feedback on stress and readiness, the underlying measurement protocols and algorithms often remain proprietary and undisclosed.

A major concern in the current ecosystem is not necessarily the technical accuracy of the sensors themselves but the transparency of the evidence base used to justify app-provided insights. Many apps cite peer-reviewed literature to establish credibility; however, it is often unclear whether these citations support the general physiological theory of HRV or if they represent specific disclosure of the app's own methodology. Without clear reporting of how metrics are calculated and what evidence supports them, users cannot distinguish between scientifically grounded tools and those using unverified metrics [10].

To address this gap, this study does not seek to validate app measurements but rather to audit the transparency and scientific attribution of the HRV app market. This study aims to describe the landscape of mobile apps that measure, analyze, and provide feedback on HRV, with a focus on how HRV is measured, analyzed, interpreted, and communicated to users. A secondary aim is to assess the transparency of these apps, including the extent to which they disclose the evidence underpinning their HRV metrics and feedback.

Methods

Study Design and Search

This study is designed as an app store review and content analysis of smartphone apps designed to measure or monitor HRV. The reporting of this study follows the Quality and Risk of Bias Checklist for Studies That Review Smartphone Applications [11] and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement guidelines [12]. A search of HRV apps was conducted in the Google Play Store (Android) and iTunes App Store (iOS) between October and November 2024 [13]. These account for approximately 99.3% of smartphone operating systems [14]. All initial searches were conducted in the Netherlands using the search terms "Heart Rate Variability" and HRV; no additional search filters were applied. Searches were conducted on both smartphone apps and web-based versions of the app store, with the exact same search terms. All identified apps were recorded in a spreadsheet. Duplicates within the same store were removed.

Selection Criteria

Initial screening was based on the App store information. Irrelevant apps were excluded at this stage. Any potentially relevant apps were retained, and eligibility was determined through website review if required. All apps were independently screened by 2 reviewers according to the inclusion and exclusion criteria outlined below. Any disagreements were solved by the research team through discussion between the 2 reviewers, with a third reviewer available in case of persistent disagreement. The reasons for exclusion were recorded. Apps were eligible for inclusion if they included functionality to record, analyze, or provide feedback on HRV and were available in English. The apps had to measure or display HRV using a smartphone-based PPG via the camera; wearable devices; or an aggregator app that imports, analyzes, or displays HRV from external wearables.

Apps were excluded if they did not feature functionality to record, analyze, or provide feedback on HRV or if they were no longer accessible during final data extraction (February 2026). Additionally, apps were excluded if they focused on nonhuman participants or were not available in English.

Data Extraction

Data were extracted from the information provided on the app store listing page, screenshots provided, and the app website where relevant. The most recent update of data extraction was done in February 2026. When details were not available or unclear, developers were contacted via email with a request to provide information. If there was no response, a second email was sent within a month. If there was still no response, the data remained unavailable. Apps were downloaded only when app developers instructed us to do so. None of the apps were used for data collection on volunteers. Data extraction was started in the Netherlands but later transitioned to Ireland. The US app stores were used in both locations to maintain consistency across the extracted data. Duplicates between stores were taken together.

We extracted app metadata (name, developer, release year, year of last update, number of downloads in February 2026, cost or payment model, platform [iOS and/or Android], primary measurement tool or aggregator app, and data storage method) and HRV measurement protocol (recording position [eg, supine, seated, and undefined], time of day, standardization procedures [eg, controlled breathing and resting conditions], and HRV measurement recording duration). With regard to data analysis and presentation, information was extracted on metrics derived (time domain; eg, root mean square of successive differences [RMSSD] and SD of normal-to-normal intervals [SDNN]; or frequency domain, such as low frequency (LF) to high frequency (HF) ratio and percentage of successive NN intervals that differ by more than 50 milliseconds) and data presentation (eg, personal trends, detailed parameters, and normal reference

XSL•FO
RenderX

ranges). Additional application functionalities included the option to add self-reported stressors (eg, stressful events, caffeine intake, comments, and workout tracker) and reminders, where available.

For all apps, data were extracted on scientific evidence cited, including whether the evidence comprised peer-reviewed literature on HRV physiology or product-specific validation studies. Furthermore, information related to the feedback provided, such as training advice, stress scores, recovery guidance, or recommendations on how to improve HRV, was recorded. A full list of extraction parameters and their descriptions is provided in Multimedia Appendix 1.

Content Analysis

Only apps with full information available on app metadata and HRV measurement and analysis, representing a transparent subset, were included in the content analysis. Apps available on both stores were reported together in the content analysis, with the first release year and the most recent year of update used for reporting purposes. The data storage statement from the Apple iTunes store was used.

Quality of Information Provided, Privacy, and Security

To evaluate the information provided to users, credibility criteria were developed from published information on the appraisal of traditional medical literature [15], authorship, attribution, disclosure, and currency. The disclosure criteria were left out; however, a criterion on privacy and security was added. Authorship examined ownership, creators, contributors, and developers, considering professional affiliations, academic credentials, and other relevant expertise. Attribution assessed whether the app or website cited scientific references to support its content. Currency of information verified the date of the most recent update for the app.

Privacy and security information were extracted based on the type of personal data collected and how the data were processed or stored. Information was extracted from both app stores, each of which uses its classification system. From the Google Play Store, apps were categorized based on whether data were only processed locally without being stored or whether they were encrypted in transit. From the iTunes App Store, data handling was categorized as one of the following: no data were stored, data were processed only locally, data were collected but were not linked to the user's identity, data were collected and linked to identifiable users, or data were used to track users across other platforms (apps and websites).

Results

App Selection

The initial search resulted in a total of 746 smartphone apps, of which 42 were within-store duplicates and 27 between-store duplicates. Following eligibility screening, 206 apps were included for initial data extraction (Figure 1). Of these, full information on app metadata and HRV measurement and analysis was obtained from a transparent subset of only 93 unique apps from 80 developers. The app names are listed in Textbox 1. The full data extraction table (N=206) is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. Workflow of the study process. *These apps had an English-language storefront but a Chinese-language developer website, which prevented complete data extraction. HRV: heart rate variability.

XSL•FO
RenderX

Textbox 1. Transparent subset of apps included in the content analysis (n=93).

- Acentas HR Monitor
- Athlytic: AI Fitness Coach
- Bevel: All-in-One Health App
- Blood Pressure app BreathNow
- BodyWave: Health tracker
- BradBeat HRV
- Breath Ball: Breathing & HRV
- BreathTuner HRV
- Camera Heart Rate Variability
- CardioBot: Heart Rate Monitor
- Complete HRV
- DailyBeat HRV
- ECG Analysis for Polar H10
- Eclipse Yourself: Health Watch
- Elite HRV: Wellness & Fitness
- ENGY – Health Monitoring based
- Firstbeat Life
- Fitbit
- FITIV Pulse Heart Rate Monitor
- FITTR HART: Smart Ring
- Flowtime: Meditation & Relax
- Garmin Connect
- Gentler Streak Workout Tracker
- Health Hive: Compare Metrics
- Healthye: Heart Rate Monitor
- Heart Analyzer: Pulse Tracker
- Heart Monitor Diary
- Heart Rate Monitor & HRV [BLE]
- Heart Rate Monitor: Cardiobyte
- Heart Rate Monitor: Pulse
- Heart Rate Variability Logger
- HeartBit: Heart Health Tracker
- HeartBreath HRV
- HeartMath
- Heartspace: HRV Training
- HeartWatch: Heart Rate Monitor
- Hera Leto
- HRV Analysis
- HRV Health
- HRV Health basic
- HRV Health Pro
- HRV Trace
- HRV Tracker for Watch

XSL•FO
RenderX

- HRV Train
- HRV Watch: Readiness Score
- HRV4Biofeedback
- HRV4Training
- iHeart HRV
- Inner Balance
- iStress: AI Stress Monitor
- Kana DL
- Kubios HRV
- Lief App
- Livity: Sleep & Health Tracker
- Me - Make health intuitive
- MindBreath – Breathe meditation
- Mindfield eSense
- Moodji: Health & Mood Tracker
- Morpheus Training
- My Autonom Health
- Neuropeak Pro
- ONVY - AI Health Coach
- Optimal HRV
- OtterLife: AI Health Tracker
- Oura
- Paced Breathing
- Polar Flow
- Professional HRV
- Pulse HRV by Camera BLE ECG
- SelfLoops HRV
- Sleep as Android Unlock
- Sleep as Android: Smart alarm
- Sleep Details
- Sonar – Health & Performance
- Stress Check: Health Monitor
- Stress Monitor – Moodpress
- Stress Monitor for Watch
- Stress Tracker: HRV Monitor
- StressEraser Pro
- StressFace: HRV Stress Tracker
- StressWatch: AI Stress Monitor
- Sweetbeat HRV
- SweetBeat HRV Classic
- Training Today
- Ultrahuman
- Visible: Pacing for illness
- Vital Science by Biostrap

XSL•FO
RenderX

- Vitalmonitor
- Wellhero: Stress & Recovery
- Welltory: Health, Heart Rate
- wHealth Dashboard
- WHOOP
- Yudemon HRV

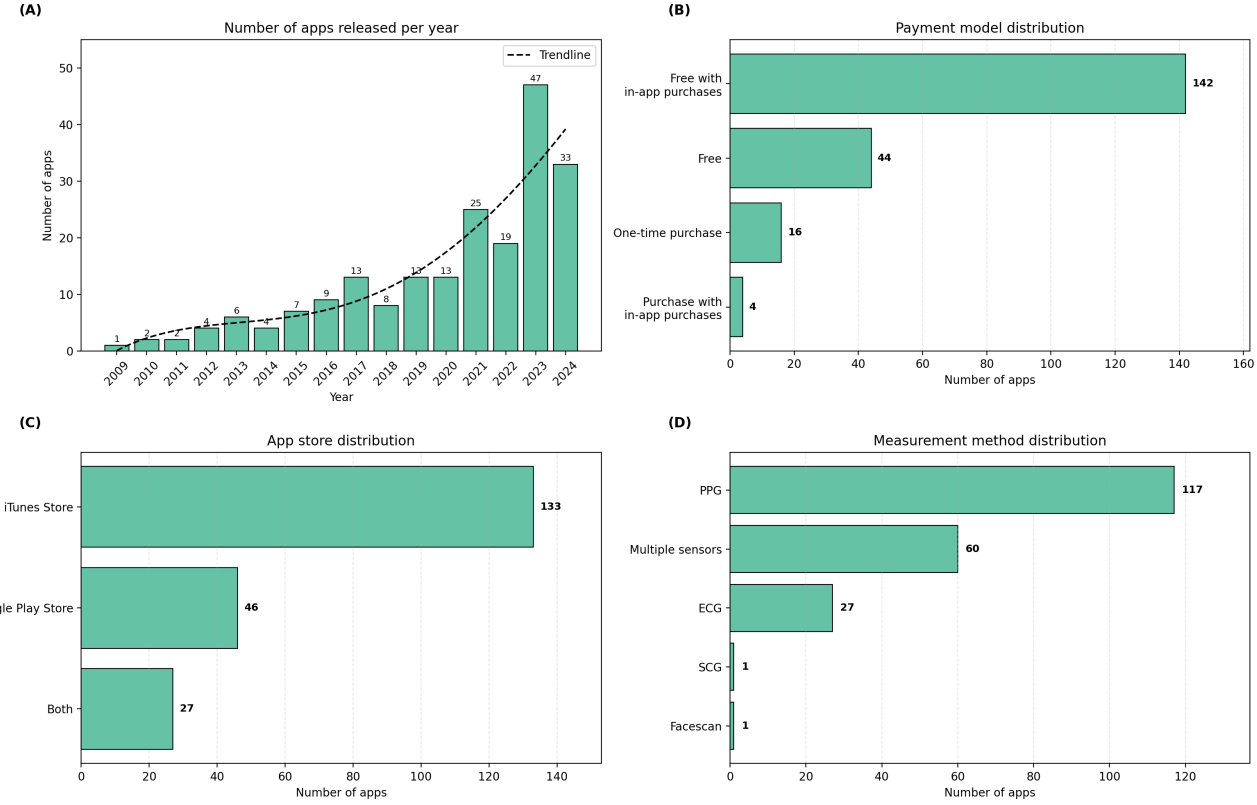
Metadata on All Apps and Missing Data

Among the 206, a total of 64.6% (n=133) originated from the Apple iTunes Store, 22.3% (n=46) were from the Google Play Store, and 13.1% (n=27) were present in both stores. In total, 63.6% (n=131) of the apps were primary measurement apps, 29.1% (n=60) were aggregators, and 7.3% (n=15) were hybrid, depending on the measurement device used. In total, 70.6% of app metadata and HRV measurement and analysis were found on the website or in the app descriptions, 5.8% were supplied by developers via email, and 23.6% of data were not provided at all.

An overview of the key characteristics of the 206 apps is presented in Figure 2. Of the 206 apps, 205 (99.5%) had information regarding device type, 203 (98.5%) had information regarding sensor type, and 190 (92.2%) described the location of the sensor. However, information about measurement description (n=127, 61.7%) and measurement frequency and duration (n=109, 52.9%) was less frequently available. The majority of apps (n=142, 68.9%) were free with in-app purchases available, and an additional 21.4% (n=44) were completely free (Figure 2). The apps supported various devices. Most apps (n=117, 56.8%) used PPG via a smartphone camera or wearable device for measurements, while 29.1% (n=60) were compatible with multiple sensors that used either PPG and/or ECG. A smaller portion (n=27, 13.1%) relied solely on wearable ECG (chest straps and wireless ECG; Figure 2D).

An exploratory analysis of app download counts indicated that among the most downloaded apps, 40% (n=4) were aggregator apps and 50% (n=5) were primary measurement apps. Most apps relied on PPG sensors (n=7, 70%). As download counts were only available as approximate numbers, these data were analyzed descriptively. Notably, only 3 of these top 10 apps were part of the transparent subset and included in the content analysis, meaning that sufficient data could be extracted from the app stores and associated websites for only these 3 apps. The top 10 apps and their characteristics are presented in Multimedia Appendix 3.

Figure 2. Key characteristics of all smartphone apps (N=206) (A) Bar chart showing the number of apps released per year, with an exponential trendline. Data for 2024 are incomplete because the search was conducted in November 2024. (B) Distribution of the different payment models. (C) Distribution of the app stores. (D) Distribution of measurement methods. ECG: electrocardiography; PPG: photoplethysmography; SCG: seismocardiography.



XSL•FO
RenderX

Content Analysis of Apps With Sufficient Information on HRV Measurement and Analysis

Of the 206 apps, full app metadata and HRV measurement and analysis data were successfully extracted for only 93 apps. These 93 apps formed the basis of the subsequent content analyses. Of the extracted app metadata and HRV measurement and analysis, 88% of the information was obtained from the app descriptions, in the app stores, or linked app websites, while the remaining 12% was collected via email. Complete information could be obtained from publicly available sources for only 49 apps. Stressors (self-reported or workout-tracking data) could be added in 52 apps, while reminder functions were available in 22 apps and unnecessary in 30 apps because measurements were continuous. The summary findings are presented in Multimedia Appendices 4 and 5.

In content analysis, PPG was the predominant sensor used for HRV measurement, with 41.9% (n=39) of the apps using PPG signals acquired from smartwatches, smartphone cameras, or PPG devices placed on the finger or ear (Table 1 summarizes sensor locations by sensor type). In total, 58.1% (n=54) of the apps were free with in-app purchases, 24.7% (n=23) were completely free, and the remaining apps required payment, with or without in-app purchases. Most apps used multiple metrics to measure HRV, and the majority used time-domain metrics to report HRV scores such as the RMSSD (n=51), the SDNN (n=48), or the percentage of adjacent normal-to-normal intervals (n=19); other apps also used time-domain metrics and frequency-domain metrics such as HF power (n=22), LF power (n=22), or the LF/HF ratio (n=15). Most measurements were user triggered, with the largest part needing to be completed first thing in the morning (Figures 3A and 3B). Continuous measurements (24 hours or overnight; Figures 3A and 3B) accounted for around 37.6% (n=35). Figures 3C and 3D show the number of apps using ultrashort measurements (<5 minutes), short-term measurements (approximately 5 minutes), and long-term measurements (>5 minutes) [1].

Table . Cross table showing the sensor used and the sensor location (n=93).

	Arm	Chest	Ear	Finger	Head	Wrist	Multiple sites
Wearable PPG[a] sensor	1	—[b]	2	7	—	27	2
Bluetooth ECG[c] HRM[d]	—	19	—	—	—	—	4
Multiple sensors possible	—	2	—	—	2	—	27

[a]PPG: photoplethysmography.

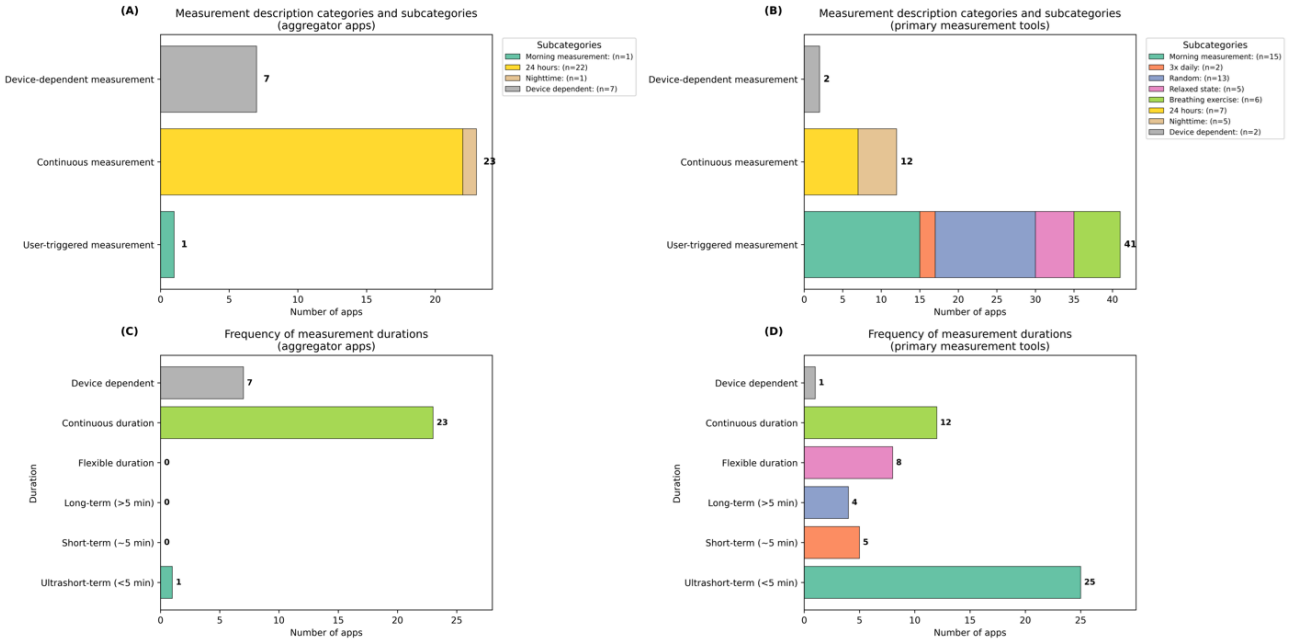
[b]Not applicable.

[c]ECG: electrocardiogram.

[d]HRM: heart rate monitor.

XSL•FO
RenderX

Figure 3. Characteristics from the content analysis stratified by aggregator apps and primary measurement apps: (A) Stacked bar chart showing the frequency of different timing and conditions of measurement for aggregator apps (n=31), (B) stacked bar chart showing the frequency of different timing and conditions of measurement for primary measurement apps (n=55), (C) measurement duration for measuring heart rate variability in aggregator apps (n=31), and (D) measurement duration for measuring heart rate variability in primary measurement apps (n=55). Device dependent indicates that the app collected data from different devices, and the measurement was dependent on the device rather than the app. Hybrid apps are not presented in the figure, as most descriptions and durations depend on the connected device.

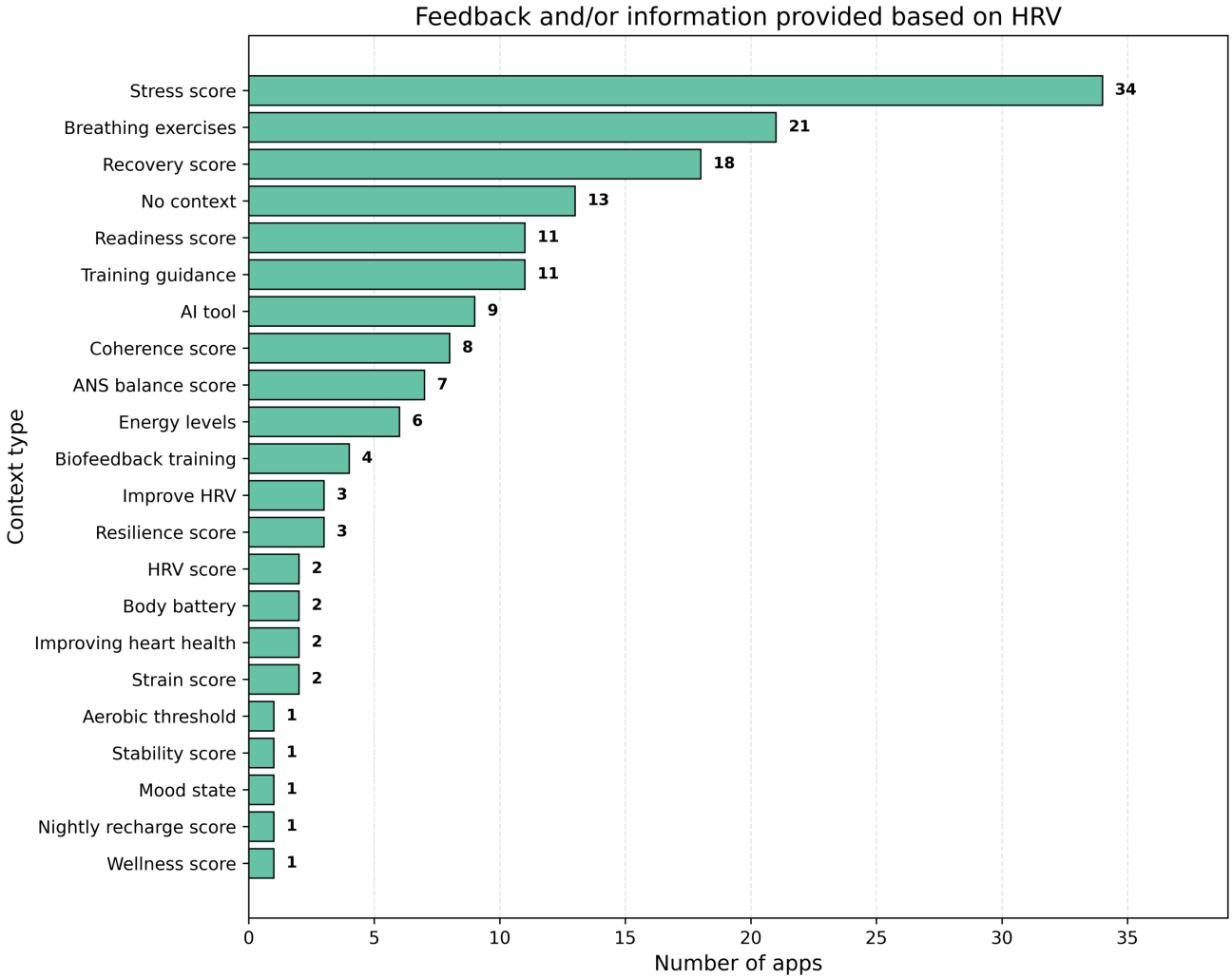


Most apps (n=76, 81.7%) provide personal trends in HRV scores. Some provide a personalized baseline range or average, calculated as a range over a period of 7 days to multiple months, or present historical data over time. More than half of the apps allowed users to log self-reported stressors (n=52, 66.7%). Scientific references were cited in most cases; the majority cited theoretical references (n=46, 58.2%), while others cited product-specific validation studies supporting the accuracy of the app's HRV measurements (n=11, 13.9%). Most (n=80, 86%) apps provided contextual HRV insights, guidance, or additional metrics such as recovery or readiness scores (Figure 4). Most apps supported data export, although some provided this functionality only upon request or suggested exporting data through Apple Health. Data storage practices varied: some apps stored no data, while others encrypted it, stored it anonymously, linked it to user identity, or used a combination of these approaches for different types of data.

Apps were stratified into primary measurement apps (n=55, 59.1%), which process raw signals, and aggregator apps (n=31, 33.3%), which use precalculated metrics from third-party hardware or data sources. Hybrid apps (n=7, 7.5%) were those whose functionality depended on the connected device used, allowing them to operate either as primary measurement or aggregator apps. Measurement characteristics also differed between categories. Continuous monitoring was most common among aggregator apps (23/31, 74.2%), while primary measurement apps more frequently used ultrashort, short, or flexible measurement durations (Multimedia Appendix 5).

XSL•FO
RenderX

Figure 4. Feedback and information provided by heart rate variability (HRV) apps after measurement. Apps could offer multiple types of feedback (eg, stress score and recovery guidance; n=93). AI: artificial intelligence; ANS: autonomic nervous system.



Quality Assessment

A clear statement of authorship was provided by 40.9% (38/93) of apps, while 61.3% (n=57) cited peer-reviewed scientific evidence to support their use or information. In terms of information currency, the majority (n=71, 76.3%) were last updated between 2025 and 2026.

Privacy and security of personal data were categorized based on app store classifications. A total of 28 (30.1%) apps collected no data and only processed data locally, 14 (15.1%) apps collected data but did not link it to user identity, 15 (16.1%) apps collected at least some data linked to user identity, 23 (24.7%) apps indicated that personal data may be collected but are encrypted in transit, and 13 (14%) apps indicated that at least some data may be used to track users across other devices. The full results of this quality assessment are available in Multimedia Appendix 6.

Discussion

Principal Findings

This study evaluated 206 mobile apps for HRV to categorize their measurement methodologies, feedback mechanisms (eg, stress and recovery scores), and scientific evidence supporting their claims. It is critical to clarify that the content analysis in this manuscript is based on apps with sufficient available data and therefore represents only 45.1% (93/206) of all identified apps, corresponding to the most “transparent” subset of the market. Conversely, most of the market (n=113, 54.9%) remains untransparent and lacks verifiable data regarding its technical protocols and scientific foundations. While the market has seen a steady increase in HRV-related app releases, this study identified a significant lack of transparency regarding how HRV was measured and the scientific evidence supporting the feedback they provided or the claims they made about their utility. Full app metadata and HRV measurement and analysis data could only be successfully extracted from 45.1% (n=93) of the eligible apps, and only 49 apps provided this information on publicly accessible channels.

Among the 93 apps included in the content analysis, the RMSSD was the primary HRV metric in both the scientific literature [16] and in these apps (n=51, 54.8% apps), followed by the SDNN (n=48, 51.6% apps). The prevalence of SDNN might be related to its role as the default metric for Apple Watch–integrated apps [17]. However, this study highlights a discrepancy in measurement validity: while Apple Watch measurements are validated when manually triggered through specific apps (eg, Breath app or Sleep Watch app), values

XSL•FO
RenderX

obtained automatically during random daily activities remain unvalidated in current literature [18-20]. Notably, the prevalence of ultrashort recordings (<5 minutes) in the analyzed apps (n=30) prioritizes users' adherence. While this is scientifically supported for RMSSD, it may compromise the reliability of metrics such as SDNN or frequency-domain powers, which typically require longer durations [1,21]. Furthermore, 86% (80/93) of apps provided contextual insights such as "readiness" or "recovery" scores. These are frequently based on proprietary algorithms that integrate multiple data points, though the underlying scientific evidence for these composite scores was often not explicitly disclosed.

Comparison With Prior Work

The dominance of PPG in 56.8% (117/206) of the apps aligns with the current trend in wearable technology. While previous research supports the validity of PPG-derived HRV under controlled, resting conditions [7,22,23], the findings of this review suggest that real-world application remains inconsistent. Some smartphone apps that use PPG via phone cameras have undergone scientific validation for HRV measurements [24,25]. Several studies report a good agreement between PPG-derived pulse rate variability and ECG-derived HRV during rest or sleep, particularly in healthy young adults. However, results are less consistent in active, real-world conditions [7], where motion and other uncontrolled conditions can reduce measurement accuracy. Only 31.2% (29/93) of the transparent subset implemented standardization protocols, such as requiring measurement during sleep without distractions (n=6, 6.5%), first thing in the morning after waking (n=17, 18.3%), or during specific resting conditions (n=6, 6.5%).

The accuracy of HRV is influenced by body position and external factors during recording [26,27]; however, only a small proportion of the apps provide specific guidance regarding posture during recording. This lack of standardization may lead to noninterchangeable data, particularly as wearables often tend to underestimate HRV values compared to the gold standard ECG [18,20].

While 61.3% (57/93) of the reviewed apps cited scientific literature to support their general methodology, many proprietary scores, such as body battery, coherence score, readiness, or stress scores, frequently lack the same level of peer-reviewed validation as the raw time-domain metrics (eg, RMSSD) [28]. This review confirms that while the hardware for PPG is increasingly accurate, the software layer providing user feedback often lacks the transparency found in clinical or research settings. Despite identifying 206 HRV apps, only 70.6% of all app metadata and HRV measurement and analysis data were found on the website or in the app descriptions, highlighting a widespread lack of transparency. Additionally, only 2 apps cited authorship, provided scientific attribution, were updated in the last 2 years (currency), and did not collect any data (privacy and security). These factors are crucial for evaluating app credibility, both in the measurement methodologies they use and in the scientific evidence supporting their claims.

Limitations

First, the terms "HRV" and "Heart Rate Variability" may have excluded wellness apps that measure HRV but do not use this terminology. Second, the search was concluded in November 2024, and although download metrics were updated in February 2026, the sample does not account for new apps released in 2025 or early 2026. Third, content analysis was restricted to 93 (45.1%) apps with full information available on app metadata and HRV measurement and analysis. Results thus represent the market's transparent subset rather than the broader market. Fourth, 1-time purchase apps were overrepresented in the content analysis compared to the full sample (13/93, 14% vs 16/206, 7.8%), while free apps with in-app purchases were underrepresented (54/93, 58.1% vs 142/206, 68.9%). This skew likely reflects the better resourcing of apps, which allowed for full details to be available and included in the analysis. Fifth, the inclusion of only English-language apps may introduce regional bias and fail to represent the global diversity of the HRV market. Finally, the assessment of privacy and security relied on self-reporting within the app stores, and technical verification (such as network traffic analysis) was not performed. Prior research has identified discrepancies between self-reported and actual data handling practices [29,30]; therefore, caution is warranted when interpreting these findings, as they reflect stated policies rather than verified compliance.

Future Directions

The findings suggest several avenues for future research and industry development. There is a critical need for the development of standardized reporting frameworks for HRV apps to ensure consistency in how metrics are calculated and displayed to users. Future studies should also focus on evaluating the clinical relevance of the feedback provided by these apps, specifically validating proprietary algorithms against established physiological markers. Additionally, improving the integration of HRV data into cardiovascular research and digital health practices will require developers to enhance data accessibility and transparency regarding their measurement protocols.

Conclusions

There are currently a large number of HRV apps available, with heterogeneity in the measurement protocols, analysis, and interpretation of HRV. While RMSSD and SDNN are the primary metrics derived, a large portion of apps lacked sufficient information to be included in the content analysis, highlighting limitations in transparency regarding the methods used. Among the included apps, feedback was often provided through proprietary scores, with limited transparency regarding the underlying algorithms or scientific evidence underpinning them.

Funding

This study received no external financial support or grants.

XSL•FO
RenderX

Data Availability

The data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

EdJ, BC, EA, and SH designed the study. EdJ conducted the app search, and screening was performed by EdJ with BC, EA, and SH each acting as second reviewer. EdJ extracted the data. EdJ, BC, EA, and SH contributed to data analysis. EDJ drafted the initial manuscript. All authors contributed to data interpretation and approved the final version of the manuscript prior to submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Feature description for data extraction.
[DOCX File, 27 KB - cardio_v10i1e84764_app1.docx]

Multimedia Appendix 2
Full dataset used in this study (N=206).
[XLSX File, 5648 KB - cardio_v10i1e84764_app2.xlsx]

Multimedia Appendix 3
Top 10 most downloaded apps in February 2026.
[DOCX File, 25 KB - cardio_v10i1e84764_app3.docx]

Multimedia Appendix 4
Results of the content analysis of apps measuring heart rate variability directly or with wearable devices (n=93).
[DOCX File, 27 KB - cardio_v10i1e84764_app4.docx]

Multimedia Appendix 5
Results of the content analysis split into aggregator apps and primary measurement apps.
[DOCX File, 28 KB - cardio_v10i1e84764_app5.docx]

Multimedia Appendix 6
Results from the quality assessment.
[DOCX File, 30 KB - cardio_v10i1e84764_app6.docx]

References

1. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258. [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
2. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;5:1040. [doi: 10.3389/fpsyg.2014.01040] [Medline: 25324790]
3. Cardoso CR, Moraes RA, Leite NC, Salles GF. Relationships between reduced heart rate variability and pre-clinical cardiovascular disease in patients with type 2 diabetes. Diabetes Res Clin Pract 2014 Oct;106(1):110-117. [doi: 10.1016/j.diabres.2014.07.005] [Medline: 25145830]
4. Euliano TY, Michalopoulos K, Singh S, et al. Photoplethysmography and heart rate variability for the diagnosis of preeclampsia. Anesth Analg 2018 Mar;126(3):913-919. [doi: 10.1213/ANE.0000000000002532] [Medline: 28991110]
5. Song T, Qu XF, Zhang YT, et al. Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC Cardiovasc Disord 2014 May 1;14:59. [doi: 10.1186/1471-2261-14-59] [Medline: 24886422]
6. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-R39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
7. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 5;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
8. Kim KB, Baek HJ. Photoplethysmography in wearable devices: a comprehensive review of technological advances, current challenges, and future directions. Electronics 2023;12(13):2923. [doi: 10.3390/electronics12132923]

XSL•FO
RenderX

9. Charlton PH, Allen J, Bailón R, et al. The 2023 wearable photoplethysmography roadmap. Physiol Meas 2023 Nov 29;44(11):111001. [doi: 10.1088/1361-6579/acead2] [Medline: 37494945]
10. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 2014;5:73. [doi: 10.3389/fphys.2014.00073] [Medline: 24578692]
11. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
13. Ryan S, Ní Chasaide N, O' Hanrahan S, Corcoran D, Caulfield B, Argent R. mHealth apps for musculoskeletal rehabilitation: systematic search in app stores and content analysis. JMIR Rehabil Assist Technol 2022 Aug 1;9(3):e34355. [doi: 10.2196/34355] [Medline: 35916688]
14. Sherif A. Market share of mobile operating systems worldwide from 1st quarter 2009 to 4th quarter 2025, by quarter. Statista. 2024. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/ [accessed 2024-11-07]
15. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor--let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
16. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput 2006 Dec;44(12):1031-1051. [doi: 10.1007/s11517-006-0119-0] [Medline: 17111118]
17. heartRateVariabilitySDNN. Apple Developer Documentation. URL: https://developer.apple.com/documentation/healthkit/hkquantitytypeidentifier/heartratevariabilitysdnn [accessed 2026-05-07]
18. O'Grady B, Lambe R, Baldwin M, Acheson T, Doherty C. The validity of Apple Watch Series 9 and Ultra 2 for serial measurements of heart rate variability and resting heart rate. Sensors (Basel) 2024 Sep 26;24(19):6220. [doi: 10.3390/s24196220] [Medline: 39409260]
19. Bonneval L, Wing D, Sharp S, et al. Validity of heart rate variability measured with Apple Watch Series 6 compared to laboratory measures. Sensors (Basel) 2025 Apr 9;25(8):2380. [doi: 10.3390/s25082380] [Medline: 40285070]
20. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):6317. [doi: 10.3390/s22166317] [Medline: 36016077]
21. Munoz ML, van Roon A, Riese H, et al. Validity of (ultra-)short recordings for heart rate variability measurements. PLoS One 2015;10(9):e0138921. [doi: 10.1371/journal.pone.0138921] [Medline: 26414314]
22. Tyapochkin K, Smorodnikova E, Pravdin P. Smartphone PPG: signal processing, quality assessment, and impact on HRV parameters. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:4237-4240. [doi: 10.1109/EMBC.2019.8856540] [Medline: 31946804]
23. Jeyhani V, Mahdiani S, Peltokangas M, Vehkaoja A. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5952-5955. [doi: 10.1109/EMBC.2015.7319747] [Medline: 26737647]
24. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography. Int J Sports Physiol Perform 2017 Nov 1;12(10):1324-1328. [doi: 10.1123/ijspp.2016-0668] [Medline: 28290720]
25. Moya-Ramon M, Mateo-March M, Peña-González I, Zabala M, Javaloyes A. Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes. Comput Methods Programs Biomed 2022 Apr;217:106696. [doi: 10.1016/j.cmpb.2022.106696] [Medline: 35172251]
26. Dobbs WC, Fedewa MV, MacDonald HV, et al. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. Sports Med 2019 Mar;49(3):417-435. [doi: 10.1007/s40279-019-01061-5] [Medline: 30706234]
27. Schmitt L, Regnard J, Millet GP. Monitoring fatigue status with HRV measures in elite athletes: an avenue beyond RMSSD? Front Physiol 2015;6:343. [doi: 10.3389/fphys.2015.00343] [Medline: 26635629]
28. Doherty C, Baldwin M, Lambe R, Burke D, Altini M. Readiness, recovery, and strain: an evaluation of composite health scores in consumer wearables. Transl Exerc Biomed 2025 Apr;2(2):128-144. [doi: 10.1515/teb-2025-0001]
29. Ali MM, Balash DG, Kodwani M, Kanich C, Aviv AJ. Honesty is the best policy: on the accuracy of Apple privacy labels compared to apps' privacy policies. arXiv. Preprint posted online on Jun 29, 2023. [doi: 10.48550/arXiv.2306.17063]
30. Rodriguez D, Jain A, Del Alamo JM, Sadeh N. Comparing privacy label disclosures of apps published in both the App Store and Google Play stores. In: 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW): IEEE; 2023. [doi: 10.1109/EuroSPW59978.2023.00022]

Abbreviations

ECG: electrocardiogram
HF: high frequency

XSL•FO
RenderX

HRV: heart rate variability
LF: low frequency
PPG: photoplethysmography
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RMSSD: root mean square of successive differences
SDNN: SD of normal-to-normal intervals

Edited by A Coristine; submitted 25.Sep.2025; peer-reviewed by CA Sanches, E Vanegas; revised version received 19.May.2026; accepted 19.May.2026; published 17.Jul.2026.

Please cite as:
de Jager E, Caulfield B, Angelidi E, Holden S
Mobile Apps for Heart Rate Variability: App Store Search and Content Analysis
JMIR Cardio 2026;10:e84764
URL: https://cardio.jmir.org/2026/1/e84764
doi:10.2196/84764

© Eline de Jager, Brian Caulfield, Evgenia Angelidi, Sinead Holden. Originally published in JMIR Cardio (https://cardio.jmir.org), 17.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone Ownership Among Cardiology Inpatients Pre–COVID-19 and Post–COVID-19 Pandemic: Secondary Analysis of Randomized Controlled Trials

Sarah Badger[1,2]; Praveen Indraratna[1,2]; Jennifer Yu[1,2]; James McVeigh[1]; Joseph Magdy[1]; Joan Li[1]; Nigel Lovell[3]; Sze-Yuan Ooi[1,2]

[1]Department of Cardiology, Prince of Wales Hospital, Barker St, Randwick, Australia

[2]School of Clinical Medicine, UNSW Sydney, Sydney, Australia

[3]School of Public Health and Community Medicine, UNSW Sydney, Sydney, Australia

Corresponding Author:
Sarah Badger
Department of Cardiology, Prince of Wales Hospital, Barker St, Randwick, Australia

Abstract

Background: Mobile health (mHealth) interventions are increasing in popularity for the management of heart failure and coronary artery disease. The use of these interventions is dependent on rates of smartphone ownership. It is estimated that approximately 90% of the Australian adult population owns a smartphone; however, international studies suggest that smartphone ownership is significantly lower in patient populations, ranging from 34% to 91%. Smartphone ownership in patients with cardiovascular disease has not previously been examined.

Objective: This study aimed to examine and compare pre–COVID-19 and post–COVID-19 pandemic smartphone ownership rates of inpatients admitted with coronary artery disease or heart failure.

Methods: Data from prescreening logs of 2 multicenter randomized controlled trials, TeleClinical Care (TCC)-Pilot and TCC-Cardiac, were reviewed. TCC-Pilot recruited patients between February 2019 and March 2020. This formed the pre–COVID-19 cohort, with 377 patients screened who lived in Sydney, had a qualifying hospital admission, and had information regarding their phone ownership status. TCC-Cardiac recruited patients from July 2021 to February 2023, with 718 patients meeting the criteria and forming the post–COVID-19 cohort. Supplemental patient demographic and medical history data were collected from the electronic medical record.

Results: In the pre–COVID-19 cohort (N=377), 194 (51.5%) patients owned smartphones, 79 (21%) owned phones that were incompatible with the mHealth intervention, and the remaining 104 (27.6%) did not own a mobile phone. Smartphone owners were predominantly male (P<.001) and more often had private health insurance (P=.002). In the post–COVID-19 cohort (N=718), 366 (51%) patients owned smartphones, 106 (14.8%) owned incompatible phones, and the remaining 246 (34.3%) did not own any mobile phone. In both cohorts, younger patients were more likely to own smartphones (P<.001). Multiple comorbidities were associated with not owning a phone.

Conclusions: Smartphone ownership accounted for just over 50% of the patients in this population. It was less common among older adults, patients with comorbidities, and those with markers of lower socioeconomic status. This needs to be considered when delivering mHealth interventions.

(JMIR Cardio 2026;10:e77141) doi:10.2196/77141

KEYWORDS

smartphone ownership; heart failure; coronary artery disease; mobile health; cardiac rehabilitation

Introduction

Cardiovascular disease is the leading cause of morbidity and mortality in Australia, accounting for 27% of deaths and 5.2% of all hospitalizations [1]. Heart failure (HF) and coronary artery disease (CAD) account for a significant proportion of hospitalizations, comprising 1.6% and 1.4%, respectively. Readmission rates of up to 20% for patients with CAD and HF in the 30 days after discharge have been observed [2-4], which is associated with increased mortality and substantial health care costs [2]. Many of these readmissions are preventable, with an audit of 3 hospitals in New South Wales finding that 27% of angina pectoris admissions and 63% of HF admissions could be prevented through effective and timely outpatient care [5]. Traditional cardiac rehabilitation (CR) reduces morbidity and cardiovascular mortality, reduces hospital admissions, and

XSL•FO
RenderX

improves quality of life [6,7]. However, CR is poorly attended, with only 20% to 30% of eligible participants attending [8-10].

The use of mobile health (mHealth) interventions to reduce readmission rates and improve CR adherence is increasing following evidence of its benefit. A meta-analysis of patients with cardiovascular disease found a significant improvement in blood pressure and HF hospitalization rates with the use of mHealth interventions [11]. A separate systematic review of 17 studies, including 11 randomized controlled trials (RCTs), found that mHealth use in HF significantly improved quality of life, and showed a positive trend in quality of self-care, number and duration of hospitalizations, and medication adherence [12]. Other benefits of mHealth include improvements in uptake, adherence, and completion of CR [13,14], reduced HF mortality [15,16], and improved daily physical activity [17] and health literacy [18].

The use of mHealth interventions is dependent upon rates of smartphone ownership among the target population; thus, an understanding of this demographic is crucial in developing interventions. It is estimated that approximately 90% of the Australian adult population owns a smartphone, with 81% using the internet daily on their smartphone [19]. However, this is not a reliable indicator of smartphone use in patients with cardiovascular disease, who are typically older than the populations surveyed by the telecommunications industry. A total of 5 studies from 2019 and 2020 identified smartphone ownership rates in patient populations to be between 34% and 91%, with 3 of those studies examining populations in the United States [20-24]. Although the sample size was considerable, ranging from 146 to 555 patients, the median ages were low at less than 49 years for patients in the emergency department and inpatients with schizophrenia, and 62 years for patients in oncology and diabetes clinics. The median age for the 5th study was not reported. The highest smartphone ownership was in patients in the emergency department, and the lowest being in an inpatient psychiatric setting. To our knowledge, smartphone ownership in the cardiac patient population is unknown. Furthermore, the impact of COVID-19 on smartphone ownership is not known.

The aim of the proposed study is to examine and compare pre–COVID-19 and post–COVID-19 pandemic smartphone ownership rates of inpatients admitted with CAD or HF and determine factors that limit smartphone ownership. This would allow an exploration of the patient groups for which mHealth interventions may be unavailable, so alternative methods, including provision of smartphones or tablets, could be considered. Due to the lockdowns and social distancing during the COVID-19 pandemic, we postulate that a greater number of people gained access to smartphones as a social outlet during periods of isolation to attend telehealth appointments and for contact tracing or vaccination requirements. We propose that those with chronic health conditions who were at higher risk of severe COVID-19 would have been more affected by the prevention methods during the COVID-19 pandemic, and thus more likely to acquire a smartphone.

Methods

This study examined prescreening data from 2 multicenter RCTs. The RCTs were conducted 15 months apart due to a New South Wales Health mandate requiring the suspension of all research during the height of the COVID-19 pandemic.

TeleClinical Care

The initial pilot study, TeleClinical Care (TCC)-Pilot (ACTRN12618001547235), recruited patients between February 2019 and March 2020 from 2 hospitals in Sydney, New South Wales [25].

Patients were eligible if they were being discharged after an admission for either HF or acute coronary syndrome, were aged ≥18 years, and owned a smartphone. Patients were excluded if they were unable or unwilling to provide informed consent, unable to operate the app or attend in-person follow-up, or traveled overseas in the first 30 days after discharge. Patients were either randomized to usual care or TCC plus usual care. The intervention arm involved measuring blood pressure, heart rate, and weight daily to input into the TCC app on their smartphone, as well as wearing a fitness band which tracked minutes of activity per day. If an input was outside of the defined limits, an alert was emailed to the monitoring team (a cardiologist and a cardiac nurse practitioner) who reviewed the alert and contacted the patient, and if deemed necessary, their general practitioner or cardiologist would alter management. The primary objective was to examine the efficacy of the TCC model compared with usual care on the incidence of 30-day hospital readmission rates.

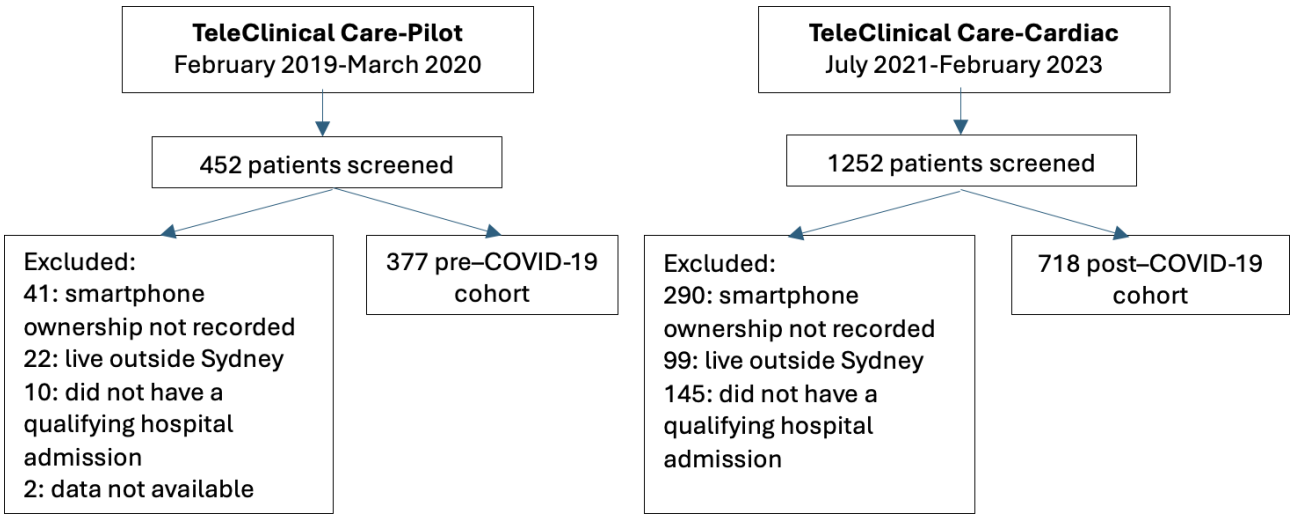
TeleClinical Care-Cardiac

The TCC-Cardiac RCT (ACTRN12621000754842) began recruitment in July 2021 from 8 public hospitals across regional and metropolitan New South Wales [26]. Patients were eligible if they were being discharged after an admission with acute myocardial infarction (AMI) or decompensated HF, provided written consent, and were aged ≥18 years. They were excluded if they had a cognitive impairment, terminal illness, limited English to operate the app or communicate with staff, enrolled in another active study, or were determined to not be able to adhere to study requirements. Patients were stratified into 1 of 3 cohorts depending on their access to technology. Cohort 1 were randomized to receive either the TCC-Cardiac model of care or usual care, cohort 2 participants were randomized to TCC-text or usual care, and cohort 3 was a registry of patients with no mobile phones receiving usual care. The TCC-Cardiac model of care collected the same data as the TCC intervention arm, as well as oxygen saturations. Additionally, the app provided medication reminders and measurement of compliance, a virtual exercise program, and education to reinforce health behaviors such as physical activity, smoking cessation, and healthy eating. The follow-up of this data was the same as the TCC intervention arm. The TCC-text intervention cohort received automated messages tailored to their diagnosis and smoking status with the aim to prompt positive behavioral change. The primary objective was the 6-month unplanned hospital readmission rate in cohort 1 patients randomized to TCC-Cardiac as an adjunct to usual care versus usual care alone.

XSL•FO
RenderX

These RCTs form the basis of this study as outlined in Figure 1.

Figure 1. Flow diagram of patient screening and cohort selection for the TeleClinical Care-Pilot and TeleClinical Care-Cardiac cohorts.



Smartphone Screening Procedure

In both trials, screening involved visually inspecting phones to confirm it was an eligible smartphone, with a smartphone being defined as a mobile phone that has internet access and is capable of downloading, installing, and operating apps. If the patient did not have a compatible smartphone, then they were either categorized as having an incompatible phone (either unable to download or operate the TCC app, or a phone that does not have internet capability) or no phone.

This Study

The TCC-Pilot data were grouped in the pre–COVID-19 cohort, while the TCC-Cardiac data were grouped in the post–COVID-19 cohort. Patients were excluded if smartphone ownership was not recorded, they did not live in Sydney, or they were not admitted secondary to HF or CAD. Data regarding smartphone ownership from TCC-Cardiac were used for this study until February 2023. Subsequently, supplemental patient demographic and medical history data were collected from the electronic medical record. The Socio-Economic Indexes for Areas Index of Relative Socio-Economic Advantage and Disadvantage was determined based on the patient's postcode. Poor mobility was a clinical judgment based on their medical history, including high falls risk or assistance required to mobilize. Patient risks and comorbidities were compared according to phone ownership using the chi-square and Kruskal-Wallis tests for categorical and continuous variables, respectively. Logistic regression was used to estimate the odds ratios for smartphone ownership adjusted for age and sex. All analyses were conducted using Stata (version 15.1; StataCorp).

Ethical Considerations

The study was reviewed and approved by the South-Eastern Sydney Local Health District Human Research Ethics Committee (2022/ETH01643). As this was a retrospective audit, consent was not required. All data were deidentified.

Results

Mobile phone ownership was assessed in 377 patients in the pre–COVID-19 cohort and 718 patients in the post–COVID-19 cohort. Demographics and comorbidities of the patients in each cohort are described in Table 1. The 2 cohorts were similar with respect to age and sex. Statistical differences were seen in some comorbidities, including primary diagnosis (P=.002), hypertension (P=.03), peripheral arterial disease (P=.01), cerebrovascular disease (P=.02), hypercholesterolemia (P<.001), and chronic kidney disease (P=.002). Smartphone ownership was similar in both cohorts, but a shift was noted from incompatible mobile phones to having no phone in the latter period (P=.04). In the pre–COVID-19 cohort, 194 (51.5%) patients owned smartphones, 79 (21.2%) owned incompatible phones, and the remaining 104 (27.6%) did not own any mobile phone. In the post–COVID-19 cohort, 366 (50.9%) patients owned smartphones, 106 (14.8%) owned incompatible phones, and the remaining 246 (34.3%) did not own any mobile phone.

XSL•FO
RenderX

Table . Comparison of the demographics and phone status of the pre–COVID-19 and post–COVID-19 cohorts.

	Pre–COVID-19 (n=377)	Post–COVID-19 (n=718)	*P* value
Age (years), median (IQR)	73 (61 - 83)	74 (64 - 83)	.18
Gender, n (%)			.24
Man	240 (63.7)	489 (68.1)	
Woman	137 (36.3)	228 (31.8)	
Nonbinary	0 (0)	1 (0.1)	
Aboriginal and/or Torres Strait Islander, n (%)	8 (2.1)	13 (1.8)	.50
Born in Australia, n (%)	171 (45.7)	324 (45.2)	.87
Private health insurance, n (%)	170 (45.2)	285 (40.2)	.03
SEIFA IRSAD[a] (decile), <9, n (%)	9 (2.4)	32 (4.5)	.36
SEIFA IRSAD[a] (decile), 9, n (%)	212 (56.2)	389 (54.2)	.36
SEIFA IRSAD[a] (decile), 10, n (%)	156 (41.4)	297 (41.4)	.36
Primary diagnosis, n (%)			.002
Acute myocardial infarction	188 (49.9)	298 (41.5)	
Unstable angina	15 (4)	14 (1.9)	
Decompensated heart failure	174 (46.2)	406 (56.5)	
Past medical history, n (%)			
Coronary artery disease	147 (39.3)	240 (33.7)	.07
Myocardial infarction	114 (30.5)	184 (25.8)	.10
Percutaneous coronary intervention	78 (20.9)	150 (21.3)	.87
Coronary artery bypass grafting	53 (14.2)	75 (10.5)	.08
Heart failure	104 (27.8)	230 (32.5)	.11
Atrial fibrillation	106 (28.3)	227 (31.9)	.22
Ventricular tachycardia	11 (3.0)	11 (1.5)	.12
Valvular disease	36 (9.6)	71 (10.1)	.81
Valvular surgery	22 (6.0)	51 (7.2)	.45
Hypertension	242 (64.2)	408 (57.3)	.03
Hypercholesterolaemia	209 (55.7)	320 (45.1)	<.001
Diabetes mellitus			.85
Type 1	7 (1.9)	13 (1.8)	
Type 2	116 (31.4)	233 (33.1)	
Chronic lung disease	50 (13.6)	86 (12.1)	.48
Chronic kidney disease	93 (24.9)	119 (17)	.002
Smoking history			.09
Current smoker	65 (17.6)	115 (17)	
Ex-smoker	116 (31.4)	173 (25.6)	
Peripheral arterial disease	43 (11.5)	48 (6.8)	.01
Cerebrovascular disease	33 (8.8)	36 (5.1)	.02
Stroke	20 (5.3)	50 (7)	.28
Transient ischemic attack	5 (1.3)	12 (1.7)	.65
Malignancy	16 (4.3)	48 (6.7)	.10

XSL•FO
RenderX

	Pre–COVID-19 (n=377)	Post–COVID-19 (n=718)	*P* value
Deep vein thrombosis or pulmonary embolus	15 (4.0)	33 (4.6)	.63
Chronic liver disease	2 (0.5)	11 (1.5)	.14
Gastroesophageal reflux disease	52 (13.9)	110 (15.4)	.49
Poor mobility	61 (16.3)	126 (17.9)	.51
Phone status, n (%)			.04
Compatible smartphone	194 (51.5)	366 (51.0)	
Incompatible phone	79 (21.0)	106 (14.8)	
No phone	104 (27.6)	246 (34.3)	

[a]SEIFA IRSAD: Socio-Economic Indexes for Areas Index of Relative Socio-Economic Advantage and Disadvantage.

In the pre–COVID-19 cohort, the median age of the participants was 73 (IQR 61 - 83) years. A total of 240 (63.7%) participants were male and 170 (45.2%) had private health insurance. Notably, private health insurance is used as a surrogate of wealth, given free health care through Medicare is available to all citizens in Australia. This is similar to the post–COVID-19 cohort, in which the average age of the participants was 74 years (IQR 64 - 83). A total of 489 (68.1%) participants were male, 13 (1.8%) identified as Aboriginal and/or Torres Strait Islander, 324 (45.2%) were born in Australia, and 285 (40.2%) had private health insurance.

In the prepandemic period, smartphone owners were predominantly male (*P*<.001) and more often had private health insurance (*P*=.002). In both cohorts, younger patients were more likely to own smartphones (*P*<.001; Table 2).

XSL•FO
RenderX

Table . Age, gender, and private health insurance stratified by phone ownership in the pre–COVID-19 and post–COVID-19 cohort, identifying the significance of phone ownership by subgroup.

Variable	Smartphone	Incompatible phone	No phone	*P* value
Pre–COVID-19				
Age (years; n=377), median (IQR)	63 (55-73)	76 (69-84)	83 (77-90)	<.001
Age (years), n (%)				<.001
<49 (n=28)	27 (96.4)	1 (3.6)	0 (0)	
50 - 59 (n=51)	39 (76.5)	7 (13.7)	5 (9.8)	
60 - 69 (n=80)	62 (77.5)	12 (15.0)	6 (7.5)	
70 - 79 (n=94)	42 (44.7)	27 (28.7)	25 (26.6)	
>80 (n=124)	24 (19.4)	32 (25.8)	68 (54.8)	
Gender, n (%)				<.001
Man (n=240)	147 (61.3)	41 (17.1)	52 (21.7)	
Woman (n=137)	47 (34.3)	38 (27.7)	52 (38.0)	
Private health insurance, n (%)				.002
Yes (n=170)	103 (60.6)	24 (14.1)	43 (25.3)	
No (n=206)	91 (44.2)	54 (26.2)	61 (29.6)	
Post–COVID-19				
Age (years; n=718), median (IQR)	68 (55-77)	74.5 (69-83)	83 (75-89)	<.001
Age (years), n (%)				<.001
<49 (n=48)	43 (89.6)	5 (10.4)	0 (0)	
50 - 59 (n=98)	84 (85.7)	9 (9.2)	5 (5.1)	
60 - 69 (n=106)	80 (75.5)	13 (12.3)	13 (12.3)	
70 - 79 (n=212)	97 (45.8)	42 (19.8)	73 (34.4)	
>80 (n=254)	62 (24.4)	37 (14.6)	155 (61.0)	
Gender, n (%)				.20
Man (n=489)	263 (53.8)	68 (13.9)	158 (32.3)	
Woman (n=228)	102 (44.7)	38 (16.7)	88 (38.6)	
Nonbinary (n=1)	1 (100)	0 (0)	0 (0)	
Private health insurance, n (%)				.27
Yes (n=285)	155 (54.4)	38 (13.3)	92 (32.3)	
No (n=424)	205 (48.3)	68 (16.0)	151 (35.6)	

In the post–COVID-19 cohort, patients who had a history of CAD (*P*=.002), HF (*P*<.001), atrial fibrillation (*P*<.001), valvular heart disease (*P*<.001), hypertension (*P*<.001), hypercholesterolaemia (*P*=.004), chronic lung disease (*P*=.02), chronic kidney disease (*P*<.001), peripheral arterial disease (*P*<.001), and gastroesophageal reflux (GORD; *P*<.001) were significantly less likely to own a smartphone (Table 3). Poor mobility was also significantly associated with having no smartphone. When this was adjusted for age and sex, patients with a history of GORD, poor mobility, valvular heart disease, atrial fibrillation, and HF were significantly less likely to own a smartphone (Figure 2).

XSL•FO
RenderX

Table . Association between comorbidities and phone ownership in the post–COVID-19 cohort, identifying the significance of phone ownership by prior diagnoses (N=718).

	Smartphone (n=366), n (%)	Incompatible phone (n=106), n (%)	No phone (n=246), n (%)	*P* value
Known CAD[a] (n=240)	100 (42)	40 (17)	100 (42)	.002
Prior AMI[b] (n=184)	82 (45)	28 (15)	74 (40)	.13
Prior PCI[c] (n=150)	60 (44)	21 (14)	69 (46)	.41
CABG[d] (n=75)	21 (29)	13 (16)	41 (55)	<.001
Heart failure (n=230)	75 (33)	38 (17)	117 (51)	<.001
Atrial fibrillation (n=227)	68 (30)	36 (16)	123 (54)	<.001
Valvular heart disease (n=71)	19 (27)	14 (20)	38 (54)	<.001
Hypertension (n=408)	180 (44)	61 (15)	167 (41)	<.001
Hypercholesterolemia (n=320)	141 (44)	54 (17)	125 (39)	.004
Diabetes mellitus (n=233)	108 (46)	32 (14)	93 (40)	.07
Chronic lung disease (n=86)	33 (38)	12 (14)	41 (48)	.02
Chronic kidney disease (n=119)	43 (36)	16 (14)	60 (50)	<.001
Peripheral arterial disease (n=48)	14 (29)	5 (10)	29 (60)	<.001
Cerebrovascular disease (n=36)	17 (47)	4 (11)	15 (42)	.61
Malignancy (n=48)	20 (42)	9 (19)	19 (40)	.41
DVT[e] or PE[f] (n=33)	11 (33)	6 (18)	16 (48)	.20
Chronic liver disease (n=11)	8 (73)	2 (18)	1 (9)	.20
GORD[g] (n=110)	35 (32)	22 (20)	53 (48)	<.001
Poor mobility (n=126)	29 (24)	23 (18)	74 (58)	<.001

[a]CAD: coronary artery disease.

[b]AMI: acute myocardial infarction.

[c]PCI: percutaneous coronary intervention.

[d]CABG: coronary artery bypass graft.

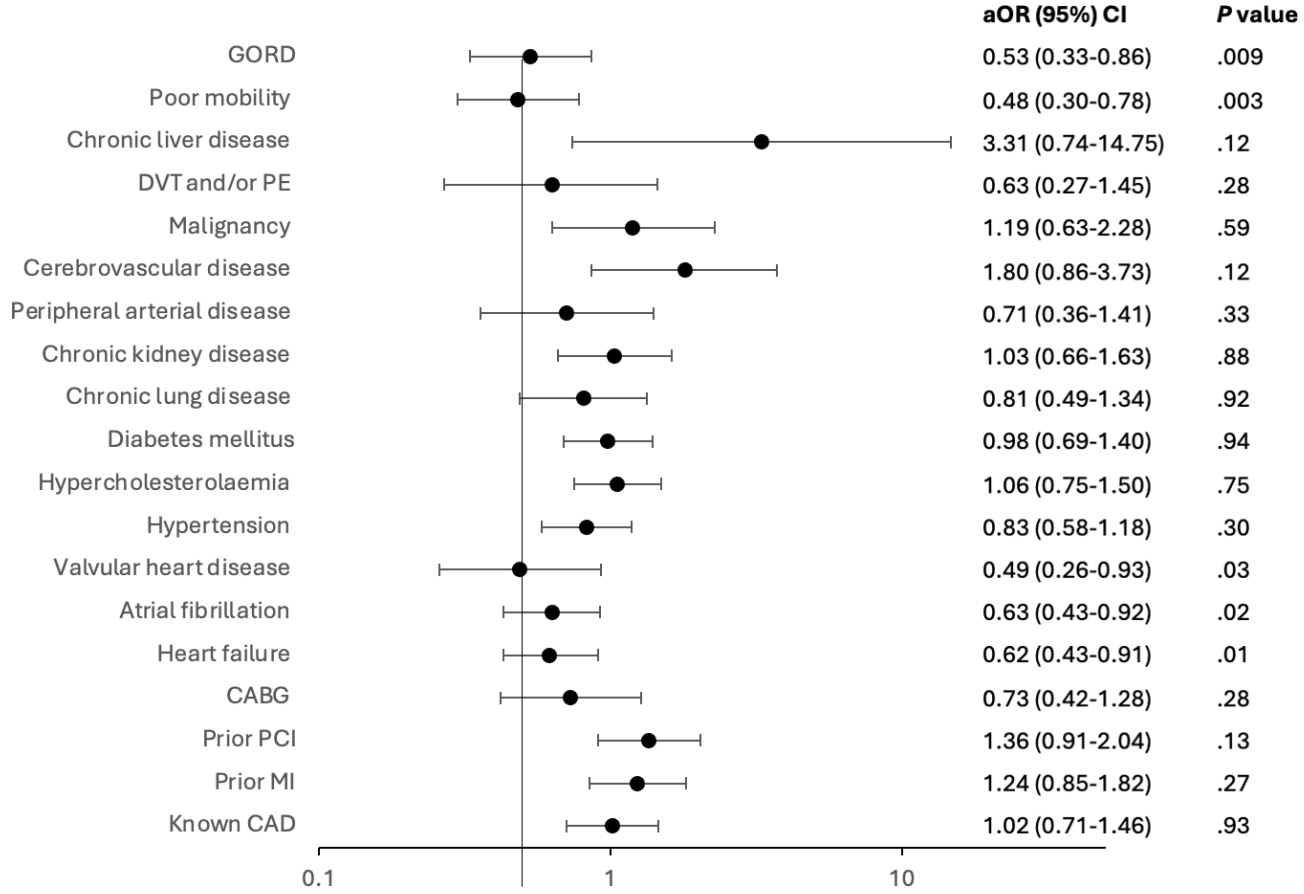
[e]DVT: deep vein thrombosis.

[f]PE: pulmonary embolism.

[g]GORD: gastroesophageal reflux disease.

XSL•FO
RenderX

Figure 2. Relationship between smartphone ownership and comorbidities when adjusted for age and sex. Adjusted odds ratio (aOR) of smartphone versus no phone for comorbidities when controlled for age and sex. CABG: coronary artery bypass graft; CAD: coronary artery disease; DVT: deep vein thrombosis; GORD: gastroesophageal reflux disease; MI: myocardial infarction; PCI: percutaneous coronary intervention; PE: pulmonary embolus.



Discussion

Principal Findings

This study is currently the largest study of smartphone ownership in an inpatient population. In this population of patients with AMI, unstable angina, and decompensated HF, smartphone ownership was just over 50%. It was less common among older adults; patients with comorbidities; and those with markers of lower socioeconomic status, such as lack of private health insurance. This study highlights that mHealth interventions could fail to address the health care disparities they set out to overcome and are likely excluding the most vulnerable population. Understanding smartphone ownership patterns is vital to tailoring mHealth interventions and improving accessibility.

Comparison With Prior Work

This study is not alone in highlighting the inequity of mHealth interventions due to low smartphone ownership rates. Similarly, in a survey of 100 outpatients with HF in the United States, 93% owned a mobile phone, with only 68% owning a smartphone [27]. In contrast, in a 2018 survey of 50 US patients (mean age 64.5, SD 8.3 years; 32% women), the study by Sohn et al [28] found that 90% of patients with HF owned a smartphone and 49% owned a tablet. However, the study population consisted of mainly White, male patients aged 50 to 80 years, which is a younger and less diverse population than our own. More recent studies have also found varying results depending on geographical location and setting examined. A study of a low socioeconomic district in South Africa investigating mHealth interventions for HIV found that of the 788 individuals surveyed, 86% owned a personal mobile phone but only 43% of these were smartphones [29]. A study of 949 dialysis patients in the United States found that 81% owned smartphones or other internet-capable devices, with 72% using the internet [30]. Rates of smartphone ownership vary significantly between patient populations; however, the trend has not been increasing in recent years when comparing studies. Our hypothesis that the COVID-19 pandemic would increase smartphone ownership appears to be false in the target population; however, the collected data present a valuable insight into smartphone ownership patterns in the surveyed population.

This is contrasted with studies looking at smartphone ownership in the general population. The Health Information National Trends Survey, a nationally representative cross-sectional survey of civilian, noninstitutionalized adults aged >18 years in the United States, found that of the 5438 responders, 87.8% were smartphone owners, 8.3% had incompatible mobile phones only, and 3.9% had no phone [31]. This cohort had equal male to female respondents with an average age of 49 years and 63.5% identifying as White. In a subset of this population who identified as having or being at risk for cardiovascular disease, there was a smaller but still high prevalence of smartphone ownership (73%) [32]. This is similar to the Pew Research Center [33] cross-sectional survey of 5626 people living in the

XSL•FO
RenderX

United States which found that 98% of people owned a mobile phone, with 91% having smartphones. Similarly, the respondents to this survey were predominantly aged <65 years and White; information on gender was not collected. Clearly, our inpatient population is not reflective of the general population in terms of age, sex, or smartphone ownership. Drawing broadly from these studies in conjunction with our own confirms that inpatients have lower rates of smartphone ownership than the general population.

A deeper analysis of smartphone ownership patterns within the patient population is required to understand the potential reach of the current and future mHealth interventions, as well as to ensure equity in these interventions. Many studies agree that there is a digital divide between younger and older adult patients, with younger people being more likely to own a smartphone [27,29-32,34]. This is especially highlighted by the Pew Research Center [33] survey which found 98% of people aged 18 to 29 years owned smartphones, but this decreased to 79% in those aged >65 years. The study by Yao et al [35] highlights in their scoping review that socially disadvantaged groups have more challenges in accessing digital health technologies, which may lead to more severe health inequities. This is supported by multiple studies showing that people who are employed, female [29], and have higher levels of education are more likely to own a smartphone [30-32,36]. Our study supports this with higher rates of private health insurance, a marker of higher socioeconomic status, being associated with smartphone ownership. Greater health literacy has also been associated with smartphone ownership [34,37]. Our study is the first to identify a detailed list of conditions associated with lower smartphone ownership; specifically, a history of GORD, poor mobility, valvular heart disease, atrial fibrillation, and HF. It is proposed that having a history of GORD is associated with not owning a smartphone, given it is known to be correlated with low socioeconomic status. However, the study by Na and Sheu [31] identified that not owning a smartphone was associated with chronic conditions such as being deaf, heart disease, lung disease, and cancer.

mHealth is a useful tool for addressing health disparities through equitable access to health information; however, there remain health inequities caused by the adoption of digital health technologies in health care. The determinants of access to mHealth are complex and often interrelated. Given that smartphones are a relatively new technology, a significant barrier is the lack of perceived usefulness and ease of use of the technology system preventing patients from prioritizing smartphone ownership [38]. There are both technology factors, such as lack of network infrastructure particularly in regional areas (eg, power, network coverage, and internet access) and the availability of smartphones, as well as patient factors (eg, age, sex, race, rurality, educational attainment, income, poor health conditions, health literacy, and education level) that contribute to not owning a smartphone.

Smartphone ownership in older adults is the most well-researched factor. The study by Wildenbos et al [39] identified 4 domains that were barriers to mHealth uptake in this population: cognition, motivation, physical ability, and perception. Learning to use mobile technology can be a challenging task due to little prior knowledge [40], lack of confidence [41], and physical impairments such as poor vision and fine motor skills [42,43]. Using mobile technology for mHealth interventions is an added complexity as the individual requires digital health literacy, a complex competency that includes health literacy, digital literacy, and the ability to make decisions on the basis of digital health data [44]. The reasons for lack of smartphone ownership are many and varied; however, it is the first barrier to accessing mHealth interventions.

Given almost half of the target population cannot access mHealth interventions in this study, strategies need to be devised to increase access to these apps to reduce hospitalization rates [11]. The study by Yao et al [35] suggests government agencies and medical institutions should provide help with resources to improve accessibility of mHealth. The aim would be to recoup the cost of these devices by avoiding hospital admissions, with mobile phones and tablets being available for AUD $200 (US $142.51) and data plans for AUD $180 (US $128.26) per year, compared with the average cost of a hospital bed in New South Wales Health of AUD $1660 (US $1182.83) per day [45], with cardiology and coronary care beds exceeding this. This suggests that the cost of a mobile phone is diminutive in comparison to readmission, a hypothesis that should be evaluated in future research. Companies should be encouraged to donate tablets and smartphones to hospitals, and software developers should be encouraged to maintain backward compatibility of their apps with older versions of devices and mobile operating systems. This would improve the accessibility to such services for socially disadvantaged groups. If a low rate of smartphone ownership is present in a population, app-based interventions may not be feasible, and alternative methodologies may be more equitable such as telephone or SMS text message–based services. Once accessibility has been overcome, the design of the intervention needs to be curated with consideration of the target group, preferably with their input during the design and implementation stages [35,46]. This will maximize accessibility and acceptability for all groups who use the app, reducing disparities. To ensure perceived usefulness, this needs to be explicitly communicated to the patient during the recruitment phase. Finally, new users need to be adequately supported when adopting novel mHealth interventions, especially when there is evidence supporting their efficacy.

Limitations

Although this is the largest study of smartphone ownership in inpatients, there are limitations. First, the study excluded patients with cognitive impairment or limited English proficiency. These patient populations often face the highest barriers to digital literacy and health care access, thus potentially underestimating the digital divide. Second, a significant number of patients were excluded from the post–COVID-19 cohort due to missing phone ownership details. The phone must have been visualized to be included in the study; this is beneficial as it ensures accuracy but also confounds the data as patients who did not bring their phone to hospital were excluded from the study. Third, it was difficult to assess the patients' sociodemographics completely given employment status was not available. The participants were primarily located at 1 local health district in a relatively affluent area in Sydney as shown by a median decile

XSL•FO
RenderX

Socio-Economic Indexes for Areas Index of Relative Socio-Economic Advantage and Disadvantage of 9 out of 10 in both cohorts. However, this tool does not account for the nuances of this area where social housing is spread among the wealthy. As such, private health insurance status was chosen to represent socioeconomic status as a reflection of discretionary spending. Fourth, the poor mobility metric was based on clinical judgment, which is subjective and may introduce bias. Fifth, as acknowledged, there are many factors affecting smartphone ownership; however, only age and sex were adjusted for. Finally, only patients with an admission diagnosis of HF, AMI, and unstable angina were included, and therefore we are unable to extrapolate to other patient populations.

Conclusions

mHealth interventions were introduced to eliminate health disparities and have the potential to improve chronic disease management. However, inequality in smartphone ownership in the highest risk comorbid patients as seen in this cohort has the potential to exacerbate rather than eliminate these disparities. Interestingly, this study showed that there was no improvement in smartphone ownership during the COVID-19 pandemic. Until smartphone ownership is commonplace in patient populations, these groups will continue to be systemically excluded. Health care providers who plan to use mHealth interventions in their practice need to ensure accessibility to all and have contingency plans if participants do not own a smartphone, such as providing a phone or tablet, or using other means to provide the same service.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

Conceptualization: PI, SO, NL
Data curation: PI, JL, JM, SB
Formal analysis: JY, PI, SB
Methodology: JY, PI, SB
Supervision: JY, PI, SO, NL
Writing – original draft: SB
Writing – review and editing: PI, JY, JM, NL
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Heart, stroke and vascular disease: Australian facts. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts/contents/about [accessed 2026-05-19]
2. Al-Omary MS, Davies AJ, Evans TJ, et al. Mortality and readmission following hospitalisation for heart failure in Australia: a systematic review and meta-analysis. Heart Lung Circ 2018 Aug;27(8):917-927. [doi: 10.1016/j.hlc.2018.01.009] [Medline: 29519691]
3. Labrosciano C, Horton D, Air T, et al. Frequency, trends and institutional variation in 30-day all-cause mortality and unplanned readmissions following hospitalisation for heart failure in Australia and New Zealand. Eur J Heart Fail 2021 Jan;23(1):31-40. [doi: 10.1002/ejhf.2030] [Medline: 33094886]
4. Rana S, Tran T, Luo W, Phung D, Kennedy RL, Venkatesh S. Predicting unplanned readmission after myocardial infarction from routinely collected administrative hospital data. Aust Health Rev 2014 Sep;38(4):377-382. [doi: 10.1071/AH14059] [Medline: 25001433]
5. Johnston J, Longman J, Ewald D, King J, Das S, Passey M. Study of potentially preventable hospitalisations (PPH) for chronic conditions: what proportion are preventable and what factors are associated with preventable PPH? BMJ Open 2020 Nov 9;10(11):e038415. [doi: 10.1136/bmjopen-2020-038415] [Medline: 33168551]
6. Anderson L, Oldridge N, Thompson DR, et al. Exercise-based cardiac rehabilitation for coronary heart disease: Cochrane systematic review and meta-analysis. J Am Coll Cardiol 2016 Jan 5;67(1):1-12. [doi: 10.1016/j.jacc.2015.10.044] [Medline: 26764059]

XSL•FO
RenderX

7. Kabboul NN, Tomlinson G, Francis TA, et al. Comparative effectiveness of the core components of cardiac rehabilitation on mortality and morbidity: a systematic review and network meta-analysis. J Clin Med 2018 Dec 4;7(12):514. [doi: 10.3390/jcm7120514] [Medline: 30518047]
8. Ades PA, Keteyian SJ, Wright JS, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the Million Hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017 Feb;92(2):234-242. [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
9. Ritchey MD, Maresh S, McNeely J, et al. Tracking cardiac rehabilitation participation and completion among Medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020 Jan;13(1):e005902. [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
10. Scott IA, Lindsay KA, Harden HE. Utilisation of outpatient cardiac rehabilitation in Queensland. Med J Aust 2003 Oct 6;179(7):341-345. [doi: 10.5694/j.1326-5377.2003.tb05588.x] [Medline: 14503895]
11. Indraratna P, Tardo D, Yu J, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e16695. [doi: 10.2196/16695] [Medline: 32628615]
12. Mouselimis D, Tsarouchas A, Vassilikos VP, et al. The role of patient-oriented mHealth interventions in improving heart failure outcomes: a systematic review of the literature. Hellenic J Cardiol 2024;77:81-92. [doi: 10.1016/j.hjc.2023.11.001] [Medline: 37926237]
13. Rosario MB, Lovell NH, Fildes J, et al. Evaluation of an mHealth-based adjunct to outpatient cardiac rehabilitation. IEEE J Biomed Health Inform 2018 Nov;22(6):1938-1948. [doi: 10.1109/JBHI.2017.2782209] [Medline: 29990228]
14. Varnfield M, Karunanithi M, Lee CK, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779. [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
15. Dendale P, De Keulenaer G, Troisfontaines P, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340. [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
16. Koehler F, Koehler K, Prescher S, et al. Mortality and morbidity 1 year after stopping a remote patient management intervention: extended follow-up results from the telemedical interventional management in patients with heart failure II (TIM-HF2) randomised trial. Lancet Digit Health 2020 Jan;2(1):e16-e24. [doi: 10.1016/S2589-7500(19)30195-5] [Medline: 33328035]
17. Felker GM, Sharma A, Mentz RJ, et al. A randomized controlled trial of mobile health intervention in patients with heart failure and diabetes. J Card Fail 2022 Nov;28(11):1575-1583. [doi: 10.1016/j.cardfail.2022.07.048] [Medline: 35882260]
18. Son YJ, Kim EY. The effectiveness and usability of a novel mobile phone-based self-care intervention for patients with heart failure: a mixed-methods pilot study. Eur J Cardiovasc Nurs 2023 Apr 12;22(3):254-263. [doi: 10.1093/eurjcn/zvac077] [Medline: 35997293]
19. Device ownership & trends. Interactive Advertising Bureau Australia. 2019. URL: https://iabaustralia.com.au/wp-content/uploads/2019/12/IAB_Device-Ownership_July-2019.pdf [accessed 2026-05-19]
20. Potdar R, Thomas A, DiMeglio M, et al. Access to internet, smartphone usage, and acceptability of mobile health technology among cancer patients. Support Care Cancer 2020 Nov;28(11):5455-5461. [doi: 10.1007/s00520-020-05393-1] [Medline: 32166381]
21. Kim E, Torous J, Horng S, et al. Mobile device ownership among emergency department patients. Int J Med Inform 2019 Jun;126:114-117. [doi: 10.1016/j.ijmedinf.2019.03.020] [Medline: 31029252]
22. Lee K, Bejerano IL, Han M, Choi HS. Willingness to use smartphone apps for lifestyle management among patients with schizophrenia. Arch Psychiatr Nurs 2019 Aug;33(4):329-336. [doi: 10.1016/j.apnu.2019.01.002] [Medline: 31280776]
23. Olamoyegun MA, Raimi TH, Ala OA, Fadare JO. Mobile phone ownership and willingness to receive mHealth services among patients with diabetes mellitus in South-West, Nigeria. Pan Afr Med J 2020;37:29. [doi: 10.11604/pamj.2020.37.29.25174] [Medline: 33062130]
24. Nathan JK, Rodoni BM, Joseph JR, Smith BW, Park P. Smartphone use and interest in a spine surgery recovery mobile application among patients in a US academic neurosurgery practice. Oper Neurosurg 2020;18(1):98-102. [doi: 10.1093/ons/opz061]
25. Indraratna P, Biswas U, McVeigh J, et al. A smartphone-based model of care to support patients with cardiac disease transitioning from hospital to the community (TeleClinical Care): pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 28;10(2):e32554. [doi: 10.2196/32554] [Medline: 35225819]
26. Yu J, Indraratna P, Briggs NE, et al. Rationale and design of the TeleClinical Care Cardiac (TCC-Cardiac) trial: a pragmatic randomized trial of adjunctive virtual models of care in the secondary prevention of cardiovascular events. Am Heart J 2025 Aug;286:56-65. [doi: 10.1016/j.ahj.2025.03.008] [Medline: 40081743]
27. Leigh JW, Gerber BS, Gans CP, Kansal MM, Kitsiou S. Smartphone ownership and interest in mobile health technologies for self-care among patients with chronic heart failure: cross-sectional survey study. JMIR Cardio 2022 Jan 14;6(1):e31982. [doi: 10.2196/31982] [Medline: 35029533]

XSL•FO
RenderX

28. Sohn A, Speier W, Lan E, et al. Assessment of heart failure patients' interest in mobile health apps for self-care: survey study. JMIR Cardio 2019 Oct 29;3(2):e14332. [doi: 10.2196/14332] [Medline: 31758788]
29. Campbell BR, Choi K, Neils MG, et al. Mobile device usage by gender among high-risk HIV individuals in a rural, resource-limited setting. Telemed J E Health 2021 Jun;27(6):615-624. [doi: 10.1089/tmj.2020.0218] [Medline: 32835638]
30. Hussein WF, Bennett PN, Pace S, et al. The mobile health readiness of people receiving in-center hemodialysis and home dialysis. Clin J Am Soc Nephrol 2020;16(1):98-106. [doi: 10.2215/CJN.11690720] [Medline: 33355235]
31. Na L, Sheu JJ. Health and information disparities among non-adopters of smartphones. Health Policy Technol 2022 Mar;11(1):100600. [doi: 10.1016/j.hlpt.2022.100600]
32. Shan R, Ding J, Plante TB, Martin SS. Mobile health access and use among individuals with or at risk for cardiovascular disease: 2018 Health Information National Trends Survey (HINTS). J Am Heart Assoc 2019 Dec 17;8(24):e014390. [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
33. Mobile fact sheet. Pew Research Center. 2025. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2026-05-19]
34. Ernsting C, Dombrowski SU, Oedekoven M, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101. [doi: 10.2196/jmir.6838] [Medline: 28381394]
35. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144. [doi: 10.2196/34144] [Medline: 35311682]
36. Bommakanti KK, Smith LL, Liu L, et al. Requiring smartphone ownership for mHealth interventions: who could be left out? BMC Public Health 2020 Jan 20;20(1):81. [doi: 10.1186/s12889-019-7892-9] [Medline: 31959145]
37. Bailey SC, O'Conor R, Bojarski EA, et al. Literacy disparities in patient access and health-related use of internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087. [doi: 10.1111/hex.12294] [Medline: 25363660]
38. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
39. Wildenbos GA, Jaspers MW, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Apr;124:68-77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
40. Wallace SE, Graham C, Saraceno A. Older adults' use of technology. Perspect Gerontol 2013 May;18(2):50-59. [doi: 10.1044/gero18.2.50]
41. Barnard Y, Bradley MD, Hodgson F, Lloyd AD. Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Comput Hum Behav 2013 Jul;29(4):1715-1724. [doi: 10.1016/j.chb.2013.02.006]
42. Fletcher J, Jensen R. Mobile health: barriers to mobile phone use in the aging population. Online J Nurs Inform 2015;19(3) [FREE Full text]
43. Searcy RP, Summapund J, Estrin D, et al. Mobile health technologies for older adults with cardiovascular disease: current evidence and future directions. Curr Geri Rep 2019 Mar;8:31-42. [doi: 10.1007/s13670-019-0270-8]
44. van Kessel R, Wong BL, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022;27:100500. [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
45. Scale of fees for hospital and other health services. NSW Government. URL: https://www1.health.nsw.gov.au/pds/Pages/doc.aspx?dn=IB2024_032 [accessed 2026-05-19]
46. Piau A, Campo E, Rumeau P, Vellas B, Nourhashémi F. Aging society and gerontechnology: a solution for an independent living? J Nutr Health Aging 2014 Jan;18(1):97-112. [doi: 10.1007/s12603-013-0356-5] [Medline: 24402399]

Abbreviations

AMI: acute myocardial infarction
CAD: coronary artery disease
CR: cardiac rehabilitation
GORD: gastroesophagealreflux disease
HF: heart failure
mHealth: mobile health
RCT: randomized controlled trial
TCC: TeleClinical Care

XSL•FO
RenderX

Edited by A Coristine; submitted 10.May.2025; peer-reviewed by L Wang; revised version received 24.Feb.2026; accepted 24.Feb.2026; published 02.Jun.2026.

Please cite as:
Badger S, Indraratna P, Yu J, McVeigh J, Magdy J, Li J, Lovell N, Ooi SY
Smartphone Ownership Among Cardiology Inpatients Pre–COVID-19 and Post–COVID-19 Pandemic: Secondary Analysis of Randomized Controlled Trials
JMIR Cardio 2026;10:e77141
URL: https://cardio.jmir.org/2026/1/e77141
doi:10.2196/77141

© Sarah Badger, Praveen Indraratna, Jennifer Yu, James McVeigh, Joseph Magdy, Joan Li, Nigel Lovell, Sze-Yuan Ooi. Originally published in JMIR Cardio (https://cardio.jmir.org), 2.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Internal Validation of the Soluble ST2, Age, and Estimated Glomerular Filtration Rate–Heart Failure Score for Predicting 30-Day Major Adverse Cardiovascular Events After Heart Failure Hospitalization in a Two-Center Vietnamese Cohort: Prospective Cohort Study

An Viet Tran[1*], MD, PhD; Nguyen Khoi Quoc La[1*], MD; Van Anh Thi Phan[1,2], MD; Bao The Nguyen[3], MSc, MD; Hoa Thai Nguyen[1], MD, PhD; Nhung Hong Thi Thai[1], MD, PhD; Tho Anh Kieu Pham[1], MD, PhD

[1]Can Tho University of Medicine and Pharmacy, 179 Nguyen Van Cu street, Tan An, Can Tho, Vietnam

[2]Thong Nhat Dong Nai General Hospital, Dong Nai, Vietnam

[3]University of Medicine and Pharmacy, Hue University, Hue, Vietnam

[*]these authors contributed equally

Corresponding Author:
Tho Anh Kieu Pham, MD, PhD
Can Tho University of Medicine and Pharmacy, 179 Nguyen Van Cu street, Tan An, Can Tho, Vietnam

Abstract

Background: Patients hospitalized for heart failure (HF) face a high-risk early postdischarge period.

Objective: We aimed to develop and internally validate the soluble ST2, age, and estimated glomerular filtration rate–heart failure (SAGE-HF) score, a simplified risk model incorporating soluble suppression of tumorigenicity 2 (sST2), age, and renal function, for predicting 30-day postdischarge major adverse cardiovascular events (MACE).

Methods: This two-center prospective cohort study included 218 patients hospitalized with acute decompensated heart failure (ADHF) and/or heart failure with reduced ejection fraction (HFrEF) and 59 non-HF controls in Vietnam (July 2025-January 2026). Serum sST2 concentrations were measured at admission using enzyme-linked immunosorbent assay and compared between the ADHF/HFrEF cohort and controls, with adjustment for clinical covariates. Among the ADHF/HFrEF cohort, the primary endpoint was 30-day MACE, defined as a composite of all-cause death or HF rehospitalization. Candidate predictors were considered based on clinical relevance, biological plausibility, and exploratory univariable associations, and prespecified parsimonious logistic regression models were evaluated using discrimination (area under the receiver operating characteristic curve [AUC]), calibration, Brier score, and Akaike information criterion. Internal validation was performed using 1000 bootstrap resamples, and model robustness was assessed using Firth penalized logistic regression. The SAGE-HF score was derived from the final model. All data were analyzed using the R environment (version 4.5.5; R Foundation for Statistical Computing).

Results: Among 277 participants, sST2 concentrations were significantly higher in patients with HF than in non-HF controls after adjustment. Among 218 patients, 47 (21.6%) experienced 30-day MACE. In multivariable analysis, age (odds ratio [OR] 1.04 per year, 95% CI 1.01 - 1.07; *P*=.02), sST2 (OR 1.06 per ng/mL, 95% CI 1.01 - 1.11; *P*=.01), and estimated glomerular filtration rate (OR 0.98 per unit, 95% CI 0.97 - 1.00; *P*=.02) were independently associated with MACE. The final model demonstrated acceptable discrimination (AUC 0.741, 95% CI 0.664 - 0.819) and good calibration, with stable performance after bootstrap validation (corrected AUC 0.725). Firth analysis yielded consistent results. The SAGE-HF score showed progressive risk stratification, with predicted 30-day MACE ranging from 7.0% to 60.1% and maintained acceptable discrimination and calibration.

Conclusions: The SAGE-HF score, incorporating age, sST2, and estimated glomerular filtration rate, demonstrated acceptable performance for predicting 30-day MACE after HF hospitalization and may support early risk stratification, pending external validation.

(JMIR Cardio 2026;10:e97879) doi:10.2196/97879

KEYWORDS

risk score; predictive model; major adverse cardiovascular events; MACE; soluble suppression of tumorigenicity 2; sST2; age; estimated glomerular filtration rate; heart failure; soluble ST2, age, and estimated glomerular filtration rate–heart failure; SAGE-HF score

XSL•FO
RenderX

Introduction

Heart failure (HF) is a major and growing global health problem, with substantial variation in epidemiology across regions and persistent gaps in data from developing countries [1,2]. After hospitalization for acute decompensation, patients enter a high-risk early postdischarge interval, often termed a "vulnerable phase," during which adverse outcomes—including rehospitalization and death—cluster soon after discharge [3]. Indeed, approximately 25% of patients hospitalized with HF are readmitted within 30 days of discharge, and mortality during this period can approach 10% [4,5]. These early events are not only clinically important but also central to health-system quality initiatives and transitional-care pathways, motivating stronger tools for early postdischarge risk identification.

Despite extensive effort, predicting early postdischarge events in HF remains challenging: readmissions arise from heterogeneous mechanisms and diagnoses, and many existing models show only modest performance or limited portability across settings [4,6]. Current guidelines emphasize guideline-directed medical therapy, careful discharge planning, and timely follow-up, but they do not provide a universally adopted, low-burden, biomarker-enhanced tool specifically optimized for very short-term postdischarge HF outcomes [7]. Identifying whether a small set of routinely available predictors, together with a biologically informative biomarker, can improve short-term risk stratification remains a practical clinical need.

Soluble suppression of tumorigenicity 2 (sST2), part of the interleukin (IL)-33/ST2 axis, is mechanistically linked to myocardial stress, inflammation, fibrosis, and remodeling; circulating sST2 acts as a decoy receptor that neutralizes IL-33 signaling [8]. In acute and chronic HF, sST2 has repeatedly demonstrated prognostic value for mortality and HF hospitalization, including additive value beyond natriuretic peptides in several studies [9-12]. However, clinically usable short-term models that incorporate sST2 remain relatively limited, especially in Southeast Asian settings. Therefore, we conducted a two-center prospective cohort study to develop and internally validate a parsimonious model and simplified point score incorporating admission sST2, namely the soluble ST2, age, and estimated glomerular filtration rate–heart failure (SAGE-HF) score, to predict 30-day postdischarge major adverse cardiovascular events (MACE) after HF hospitalization.

Methods

Study Design and Population

This prospective observational study was conducted at Thong Nhat Dong Nai General Hospital and Can Tho Central General Hospital between July 2025 and January 2026. The study population comprised patients hospitalized with acute decompensated heart failure (ADHF) and/or heart failure with reduced ejection fraction (HFrEF) who met the prespecified eligibility criteria. Patients were eligible if they were admitted with ADHF regardless of left ventricular ejection fraction (LVEF) phenotype and/or had HFrEF documented during the index hospitalization. Underlying HF was diagnosed by the treating cardiologists based on compatible symptoms and signs, echocardiographic assessment, and routine clinical investigations during the index hospitalization. ADHF was recorded when patients were admitted because of new or worsening symptoms and signs of HF requiring inpatient treatment. LVEF was assessed by transthoracic echocardiography during the index hospitalization and was used to classify HF phenotype as HFrEF, heart failure with mildly reduced ejection fraction (HFmrEF), or heart failure with preserved ejection fraction (HFpEF). For this study, HFrEF was defined as LVEF ≤40%, HFmrEF as LVEF 41% to 49%, and HFpEF as LVEF ≥50%. Patients with HFmrEF and HFpEF were included only if they were hospitalized with ADHF during the index hospitalization. Therefore, ADHF and HFrEF were not considered mutually exclusive categories. ADHF was considered a clinical presentation, whereas LVEF phenotype was considered an echocardiographic classification. Patients with end-stage chronic kidney disease (estimated glomerular filtration rate [eGFR]<15 mL/min/1.73 m^2) or receiving dialysis, those with conditions known to affect sST2 levels, including systemic diseases such as lupus, rheumatoid arthritis, and chronic obstructive pulmonary disease, as well as those lost to follow-up, were excluded. Hospital admission and discharge decisions were made by the treating physicians according to routine clinical practice, and patients were discharged when they were considered clinically stable and suitable for outpatient follow-up. In addition, a non-HF comparison group was recruited for exploratory comparison of baseline characteristics and serum sST2 concentrations. These participants had no known history of HF based on medical record review and clinical history. They underwent clinical assessment to exclude HF, with routine electrocardiogram assessment as part of hospital evaluation. The absence of HF was further supported by N-terminal pro-B-type natriuretic peptide testing and transthoracic echocardiography. A total of 218 patients with ADHF and/or HFrEF and 59 non-HF controls were included in the final analysis.

Ethical Considerations

The study protocol was reviewed and approved by the Biomedical Research Ethics Committee of Can Tho University of Medicine and Pharmacy, Vietnam, for both participating centers. Ethical approval was granted for the study conducted at Can Tho Central General Hospital under approval number 25.326.HV/PCT-HĐĐĐ, dated June 30, 2025, and for the study conducted at Thong Nhat Dong Nai General Hospital under approval number 25.274.HV/PCT-HĐĐĐ, dated June 30, 2025. The study was conducted in accordance with the ethical principles for biomedical research involving human participants. Written informed consent was obtained from all participants before enrollment. All patient data were coded and kept confidential, and the results were reported only in aggregate form without identifying individual participants. No financial compensation or payment was provided to participants. The manuscript and supplementary materials do not contain any identifiable individual participant images or personal information.

XSL•FO
RenderX

Outcomes Assessment

The primary outcome was the 30-day postdischarge MACE. For the main outcome analysis, MACE was defined as a patient-level composite end point of all-cause death or rehospitalization for HF during follow-up [13]. During the 30-day postdischarge period, outcome data were initially ascertained by an independent assessor through telephone contact with patients or their family members. This assessor was not involved in sST2 measurement, statistical analysis, or model development and was not provided with baseline sST2 results. Reported events were subsequently verified through a review of hospital and electronic medical records. Rehospitalization for HF was defined as hospital admission due to worsening symptoms or signs of HF requiring inpatient treatment. Deaths during follow-up were identified through telephone follow-up and verified using available medical records whenever possible. Patients were subsequently classified into the MACE and non-MACE groups according to the occurrence of the composite end point within 30 days after discharge. Each patient contributed only once to the primary MACE end point. If both HF rehospitalization and death occurred within 30 days, the patient was counted once in the composite end point and was not double-counted.

sST2 Concentration Measurements

Serum sST2 concentration was measured using the Elabscience Human sST2 (Soluble ST2) enzyme-linked immunosorbent assay (ELISA) Kit (catalog number E-EL-H6082), a sandwich ELISA assay intended for research use, in accordance with the manufacturer's instructions. According to the manufacturer's documentation, the assay sensitivity was 0.19 ng/mL, and the detection range was 0.31 - 20 ng/mL. Manufacturer-reported intra-assay coefficients of variation were 4.43%, 4.36%, and 6.31% across low, medium, and high concentration samples, respectively, and inter-assay coefficients of variation were 8.22%, 6.76%, and 6.58%, respectively. Blood samples were collected at hospital admission during the index hospitalization for patients with HF and at enrollment for non-HF comparison participants. Serum samples were processed according to routine laboratory procedures and the manufacturer's instructions. Samples with optical density values exceeding the upper limit of the standard curve were not extrapolated. Instead, these samples were initially diluted 1:2 using the kit-provided Reference Standard & Sample Diluent and reassayed. If the measured concentration remained above the standard curve range, additional serial dilution was performed as needed until the measured concentration fell within the quantifiable range. Final concentrations were calculated by multiplying the measured concentration by the corresponding total dilution factor. Samples and standards were analyzed in duplicate, optical density was measured at 450±2 nm, and sST2 concentrations were calculated from a 4-parameter logistic standard curve. sST2 values were recorded in nanograms per milliliter and analyzed as a continuous variable to evaluate their association with 30-day postdischarge MACE.

Other Variables

Baseline variables were collected at admission and included demographic characteristics, clinical profile, medical history, and laboratory parameters relevant to HF prognosis. Age was recorded in years, and BMI was calculated as weight in kilograms divided by height in meters squared. Functional status was assessed according to the New York Heart Association classification [7]. Smoking was defined as a patient who had smoked at least 100 cigarettes during their lifetime and had smoked within 30 days before hospitalization [14]. Alcohol consumption was defined as any self-reported intake of alcoholic beverages, including beer, wine, or spirits, during the 12 months before hospitalization [15]. Comorbidities and clinical history, including hypertension, diabetes mellitus, dyslipidemia, prior myocardial infarction, stroke, atrial fibrillation, smoking, and alcohol use, were obtained from medical records and direct patient interview. LVEF was assessed by transthoracic echocardiography during the index hospitalization, preferentially using the biplane Simpson's method when image quality was adequate [16]. Serum sodium, urea, and creatinine were measured using routine biochemical assays at admission. eGFR was calculated from serum creatinine using the Chronic Kidney Disease Epidemiology Collaboration equation [17].

Data Analysis

All statistical analyses were performed in R (R Foundation for Statistical Computing). Data management and variable transformation were conducted primarily with dplyr, while tabular regression outputs were organized with broom and tibble. Continuous variables were summarized as mean (SD) or median (IQR), as appropriate, and categorical variables as frequencies and percentages. Comparisons between patients with HF and controls were performed using 2-tailed independent-samples *t* tests or Mann-Whitney *U* tests for continuous variables, and chi-square tests or Fisher exact tests for categorical variables, as appropriate. For sST2 comparisons between groups, adjusted analyses were performed using analysis of covariance controlling for confounders. In addition, baseline characteristics were compared between the MACE and non-MACE groups using the 2-tailed independent-samples *t* test for continuous variables and the chi-square test or Fisher exact test, or Fisher Freeman Halton exact test for categorical variables, as appropriate. Crude associations with MACE were explored using univariable logistic regression, and crude odds ratios (ORs) with 95% CIs were reported. Candidate predictors for multivariable modeling were chosen based on clinical plausibility together with univariable screening, with preference given to variables with $P<.20$ while avoiding the redundant representation of the same biological domain and limiting model complexity in view of the available number of events. A 2-sided P value <.05 was considered statistically significant for the final multivariable analyses [18,19]. Model development and reporting were guided by transparent prediction-model reporting principles, including TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) and TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis plus Artificial Intelligence) recommendations [20,21]. With 47 events, parsimonious models were limited to 3 to 4 predictors, corresponding to approximately 15.7 to 11.8 events per variable, which is consistent with conventional events per variable guidance for logistic regression [20,21]. Sample size

XSL•FO
RenderX

adequacy for prediction-model development was additionally assessed using the pmsampsize framework for binary outcomes proposed by Riley et al [22].

To reduce the risk of overfitting, model development followed a prespecified parsimonious candidate-model strategy rather than unrestricted stepwise selection. All candidate models were fitted on the same complete-case dataset to ensure a fair comparison of model performance. There were no missing data for the variables included in the candidate prediction models. All 218 patients had available sST2 measurements, baseline clinical and laboratory variables, and 30-day outcome status. Based on the clinically prioritized core predictors age and sST2, 11 prespecified logistic regression models were examined. Final model selection was based on a joint assessment of optimism-corrected discrimination, calibration, overall fit, parsimony, and clinical interpretability. When candidate models showed comparable optimism-corrected discrimination, preference was given to the model with fewer predictors, a more favorable calibration slope, and greater feasibility for conversion into a simple point-based score. Apparent discrimination for each candidate model was quantified with the area under the receiver operating characteristic curve (AUC) using pROC, with 95% CIs estimated by the DeLong method. Overall fit was additionally assessed with the Akaike information criterion (AIC) and Brier score. Although hypertension showed a signal in univariable analysis, categorical variables with sparse subgroup counts or unstable estimates, particularly hypertension and New York Heart Association class, were not prioritized for the prespecified parsimonious multivariable models. This decision was made because some categories contained relatively small numbers of patients, resulting in wide CIs and reduced estimate stability, with the potential to increase model complexity and overfitting without clear incremental predictive value [18].

Shortlisted candidate models were then subjected to internal validation using bootstrap resampling with the rms package. Apparent and optimism-corrected performance were evaluated with 1000 bootstrap resamples, focusing on AUC, Brier score, calibration intercept, and calibration slope. The final model was selected according to the prespecified rule described above, with priority given to comparable optimism-corrected discrimination, calibration slope, parsimony, and clinical interpretability. The final coefficients and adjusted ORs were then estimated using standard multivariable logistic regression with the glm function under a binomial link. The assessment of calibration was emphasized because adequate calibration is essential for the clinical usefulness of risk prediction models [18,19].

A sensitivity analysis was performed using Firth penalized logistic regression with the logistf package to assess the robustness of coefficient estimates and inference under small-sample conditions and potential sparse-data bias [23,24]. Discrimination of the final logistic model was evaluated by AUC with a 95% CI using pROC. Calibration was examined both numerically and graphically using rms, including the calibration intercept and slope as summary indices and bootstrap-based calibration plots for visual assessment of agreement between predicted and observed risk. Figures were generated with ggplot2; when a smoothed receiver operating characteristic curve was displayed for visual presentation, the reported AUC and 95% CI were still calculated from the original empirical receiver operating characteristic curve.

A simplified point-based score was subsequently derived from the final parsimonious model. First, uniform shrinkage was applied to the regression coefficients using the bootstrap-corrected calibration slope from the final model. The shrunken coefficients were then converted into a pragmatic integer score using clinically interpretable reference values and increments. For each predictor, the shrunken coefficient was first rescaled according to its clinically defined increment: 5 years for age above 60 years, 5 ng/mL for sST2 above 20 ng/mL, and 10 mL/min/1.73 m² for eGFR below 60 mL/min/1.73 m². Let w_j denote the rescaled shrunken coefficient for predictor j. Integer points were assigned by dividing each w_j by the smallest absolute rescaled coefficient and rounding to the nearest integer: $\text{Points}_j = \text{round}\ (w_j/\min(|w|))$. Using this approach, age, sST2, and eGFR were assigned 1, 2, and 1 points, respectively. The total score was obtained by summing the component points across predictors. Negative component points were not allowed. Scores above the maximum observed value in the derivation cohort were not used to generate separate risk estimates and should be interpreted as ≥12 until externally validated. A secondary logistic regression model with total score as the sole predictor was then fitted to translate score values into estimated probabilities of MACE. In this score-only logistic model, the intercept was −2.592, and the coefficient for each 1-point increase in the SAGE-HF score was 0.250, corresponding to an OR of 1.284 per 1-point increase. The predicted probability was calculated as logit(p)=–2.592 + 0.250 × SAGE-HF score. The discrimination of the simplified score was assessed by AUC with 95% CI using pROC, and its calibration was evaluated with Brier score, calibration intercept, calibration slope, and calibration plots using rms [18,19]. Decision curve analysis was performed to evaluate the clinical usefulness of the SAGE-HF score by estimating net benefit across threshold probabilities from 0.05 to 0.50 and comparing the score with treat-all and treat-none strategies [25].

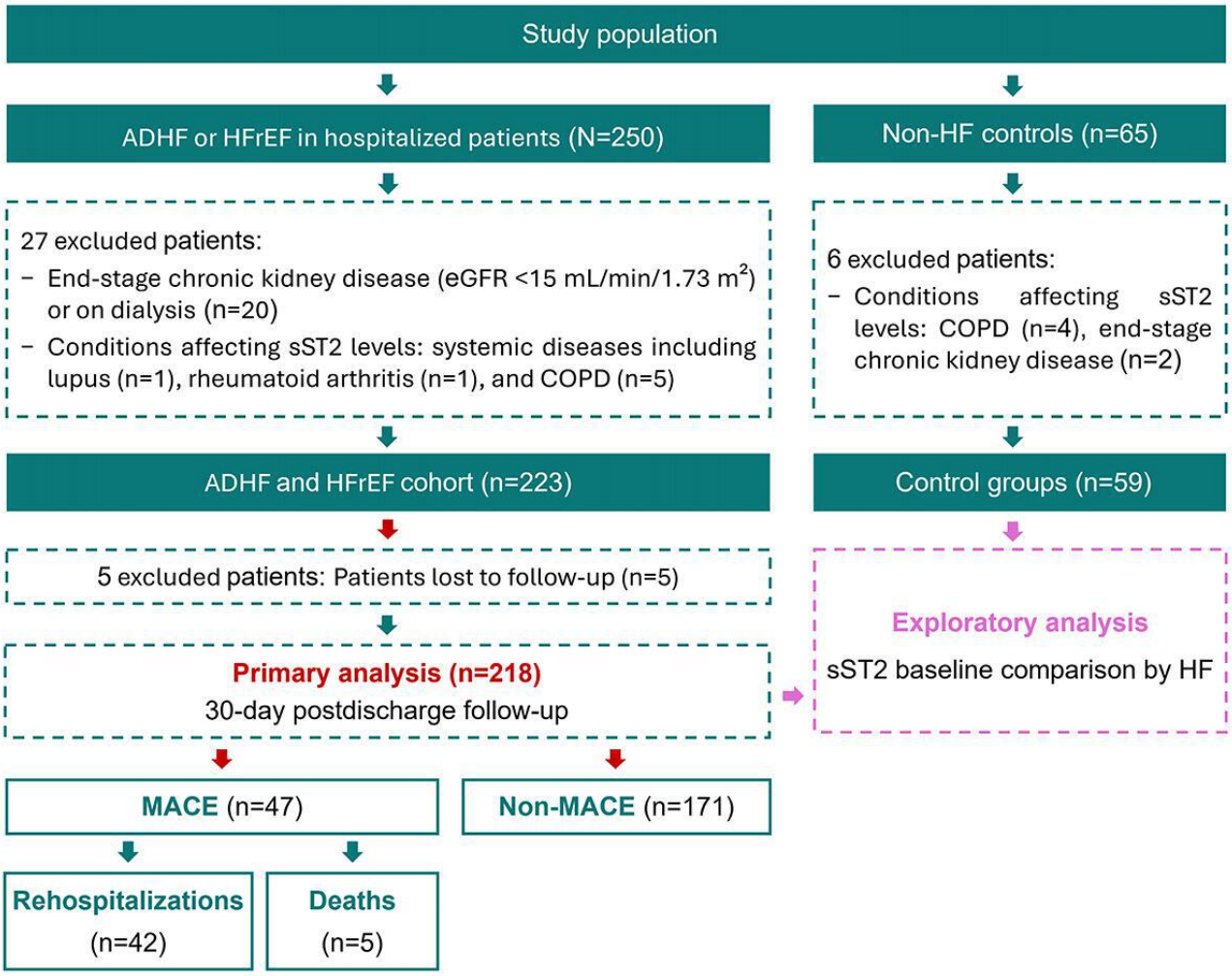
Results

A total of 218 patients with ADHF and/or HFrEF and 59 non-HF controls were included in the final analysis. The participant selection flow is presented in Figure 1. Serum sST2 concentrations were significantly higher in the HF group compared with the non-HF group (median 24.2, IQR 17.0-30.4 ng/mL vs 15.9, IQR 11.7-23.2 ng/mL; P<.001). This difference remained statistically significant after adjustment for potential confounders, with both unadjusted and adjusted analyses demonstrating consistent results (Table S1 in Multimedia Appendix 1). Among the 218 patients with HF, 47 patients experienced 30-day postdischarge MACE, including 42 rehospitalizations for HF and 5 deaths. The baseline characteristics of the overall study population and their comparisons between the MACE and non-MACE groups are presented in Table 1. Patients in the MACE group were significantly older than those in the non-MACE group (mean

XSL•FO
RenderX

74.9, SD 13.0 y vs 68.3, SD 12.1 y; *P*=.001). Additionally, patients with MACE had higher median sST2 concentrations (median 27.6, IQR 23.3-30.6 ng/mL vs 22.6, IQR 15.9-30.2 ng/mL; *P*=.006), higher urea levels (median 10.1, IQR 6.4-13.3 mmol/L vs 6.7, IQR 5.0-9.2 mmol/L; *P*=.001), and lower eGFR (median 44.0, IQR 30.7-61.9 mL/min/1.73 m² vs 62.2, IQR 45.5-84.7 mL/min/1.73 m²; *P*<.001). Serum sodium was also lower in the MACE group (median 134.0, IQR 131.6-138.5 mmol/L vs 136.0, IQR 134.0-139.0 mmol/L; *P*=.02)

Figure 1. Flow diagram of patients hospitalized with acute decompensated heart failure (ADHF) and/or heart failure with reduced ejection fraction (HFrEF) and non–heart failure (HF) controls. COPD: chronic obstructive pulmonary disease; eGFR: estimated glomerular filtration rate; MACE: major adverse cardiovascular event; sST2: soluble suppression of tumorigenicity 2.



XSL•FO
RenderX

Table . Baseline characteristics of the study population overall and stratified by 30-day postdischarge major adverse cardiovascular events (MACE).

Characteristics	MACE (n=47)	Non-MACE (n=171)	Total (n=218)	*P* value
Age (y), mean (SD)	74.9 (13.0)	68.3 (12.1)	69.7 (12.6)	.001[a]
Female, n (%)	18 (38.3)	79 (46.2)	97 (44.5)	.41[b]
BMI (kg/m^2), median (IQR)	21.3 (19.7-22.5)	22.4 (19.7-23.8)	21.8 (19.7-23.6)	.13[c]
NYHA[d] classification, n (%)				.18[e]
NYHA II	3 (6.4)	28 (16.4)	31 (14.3)	
NYHA III	20 (42.6)	82 (48.0)	102 (46.8)	
NYHA IV	24 (51.1)	61 (35.7)	85 (39.0)	
Hypertension, n (%)	40 (85.1)	162 (94.7)	202 (92.7)	.03[b]
Diabetes, n (%)	16 (34.0)	67 (39.2)	83 (38.1)	.52[b]
Dyslipidemia, n (%)	16 (34.0)	64 (37.4)	80 (36.7)	.73[b]
Prior myocardial infarction, n (%)	25 (53.2)	88 (51.5)	113 (51.8)	.83[b]
Stroke, n (%)	2 (4.3)	14 (8.2)	16 (7.3)	.36[f]
Atrial fibrillation, n (%)	10 (21.3)	31 (18.1)	41 (18.8)	.63[b]
Smoking, n (%)	11 (23.4)	47 (27.5)	58 (26.6)	.58[b]
Drinking, n (%)	15 (31.9)	63 (36.8)	78 (35.8)	.53[b]
LVEF[g] (%), median (IQR)	39.0 (32.5-56.5)	36.0 (29.5-45.0)	36.0 (30.0-48.8)	.17[c]
HF[h] phenotype, n (%)				.32[e]
HFrEF[i]	28 (59.6)	119 (69.6)	147 (67.4)	
HFmrEF[j]	4 (8.5)	15 (8.8)	19 (8.7)	
HFpEF[k]	15 (31.9)	37 (21.6)	52 (23.9)	
ADHF[l], n (%)	22 (46.8)	57 (33.3)	79 (36.2)	.09[b]
sST2[m] (ng/mL), median (IQR)	27.6 (23.3-30.6)	22.6 (15.9-30.2)	24.2 (17.0-30.4)	.006[c]
Serum sodium (mmol/L), median (IQR)	134.0 (131.6-138.5)	136.0 (134.0-139.0)	136.0 (133.0-139.0)	.02[c]
Serum urea (mmol/L), median (IQR)	10.1 (6.4-13.3)	6.7 (5.0-9.2)	7.2 (5.0-10.0)	.001[c]
Serum creatinine (µmol/L), median (IQR)	123.0 (78.5-177.0)	94.0 (71.5-124.0)	98.0 (75.0-129.8)	.03[c]
eGFR[n] (mL/min/1.73 m^2), median (IQR)	44.0 (30.7-61.9)	62.2 (45.5-84.7)	57.0 (40.9-82.5)	<.001[c]

[a]Two-tailed independent sample *t* test.

[b]Chi-square test.

[c]Mann-Whitney *U* test.

[d]NYHA: New York Heart Association. No enrolled patients were classified as NYHA class I.

[e]Fisher-Freeman-Halton exact test.

[f]Fisher exact test.

[g]LVEF: left ventricular ejection fraction.

[h]HF: heart failure.

[i]HFrEF: heart failure with reduced ejection fraction.

[j]HFmrEF: heart failure with mildly reduced ejection fraction.

[k]HFpEF: heart failure with preserved ejection fraction.

XSL•FO
RenderX

[l]ADHF: acute decompensated heart failure.

[m]sST2: soluble suppression of tumorigenicity 2.

[n]eGFR: estimated glomerular filtration rate.

Among the prespecified candidate models, discrimination ranged from an AUC of 0.714 to 0.747. The model including age, sST2, sodium, and eGFR (M8) showed the highest apparent discrimination (AUC 0.747, 95% CI 0.670 - 0.824) and the lowest AIC (207) and Brier score (0.145), whereas the model including age, sST2, and eGFR (M3) demonstrated comparable performance (AUC 0.741, 95% CI 0.664 - 0.819) with only minimal differences in AIC and Brier score (Table S3 in Multimedia Appendix 1). After bootstrap validation with 1000 resamples, the M3 model showed stable performance, with an optimism-corrected AUC of 0.725, a Brier score of 0.155, a calibration intercept of 0.009, and a calibration slope of 0.930. Although M8 had slightly better apparent performance, its optimism-corrected AUC was lower than that of M3 (0.722 vs 0.725), and its calibration slope was less favorable (0.881 vs 0.930; Table S4 in Multimedia Appendix 1). Given the minimal incremental gain in discrimination and overall fit with M8, the M3 model was selected as the final model to preserve parsimony. This choice reduced the number of predictors while maintaining comparable predictive performance, supporting model stability in the context of a limited number of events and aligning with recommended events-per-variable considerations. In addition, the exclusion of sodium avoided potential redundancy with other renal and hemodynamic markers, further enhancing model interpretability and clinical applicability. In the final multivariable logistic regression model (Table 2), age (OR 1.04 per year, 95% CI 1.01 - 1.07; *P*=.02), sST2 (OR 1.06 per unit, 95% CI 1.01 - 1.11; *P*=.01), and eGFR (OR 0.98 per unit, 95% CI 0.97 - 1.00; *P*=.02) were independently associated with 30-day MACE. Firth penalized logistic regression yielded similar results, confirming the robustness of the associations for age (OR 1.036, 95% CI 1.01 - 1.07; *P*=.02), sST2 (OR 1.055, 95% CI 1.01 - 1.10; *P*=.01), and eGFR (OR 0.982, 95% CI 0.97 - 1.00; *P*=.02; Table 3).

The final multivariable model demonstrated acceptable discrimination (AUC=0.741, 95% CI 0.664 - 0.819) and calibration for predicting 30-day MACE (Figure 2).

Table . Final multivariable logistic regression model for predicting major adverse cardiovascular events (MACE).

Predictor	Unit of comparison	Coefficient	Adjusted OR[a] (95% CI)	*P* value
Age	+1 year	0.036	1.04 (1.01 - 1.07)	.02
sST2[b]	+1 ng/mL	0.055	1.06 (1.01 - 1.11)	.01
eGFR[c]	+1 mL/min/1.73 m²	−0.018	0.98 (0.97 - 1.00)	.02

[a]OR: odds ratio.

[b]sST2: soluble suppression of tumorigenicity 2.

[c]eGFR: estimated glomerular filtration rate.

Table . Firth penalized logistic regression sensitivity analysis.

Predictor	Unit of comparison	Coefficient	Adjusted OR[a] (95% CI)	*P* value
Age	+1 year	0.035	1.036 (1.01 - 1.07)	.02
sST2[b]	+1 ng/mL	0.054	1.055 (1.01 - 1.10)	.01
eGFR[c]	+1 mL/min/1.73 m²	−0.018	0.982 (0.97 - 1.00)	.02

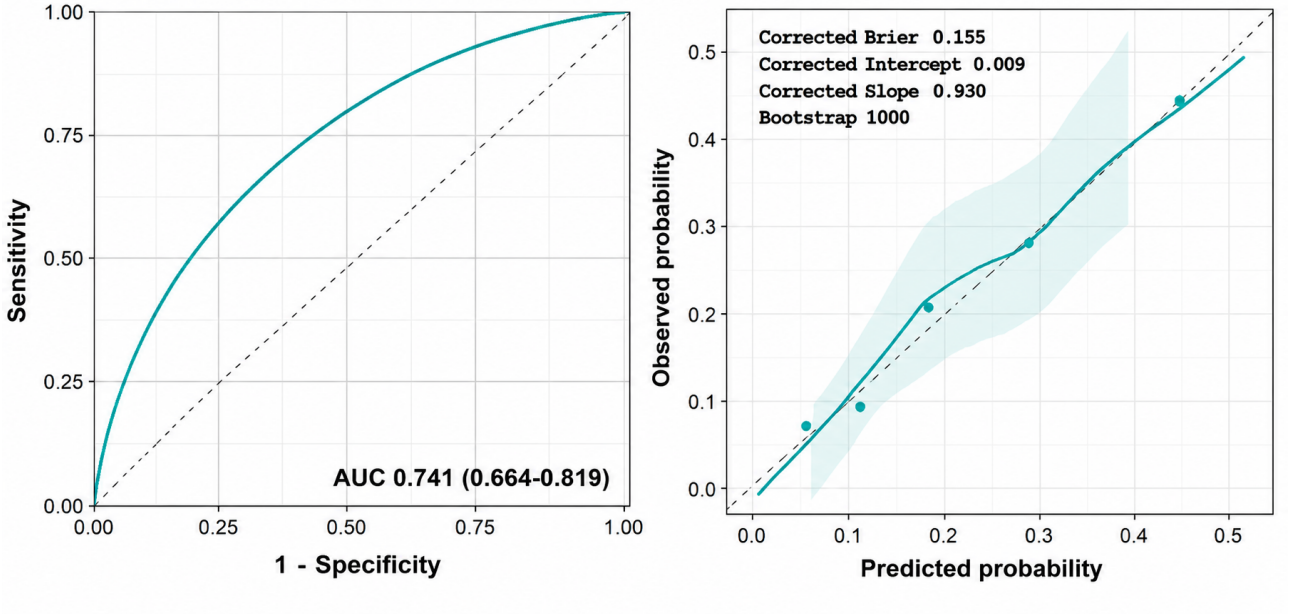
[a]OR: odds ratio.

[b]sST2: soluble suppression of tumorigenicity 2.

[c]eGFR: estimated glomerular filtration rate.

XSL•FO
RenderX

Figure 2. Performance of final multivariable logistic regression model. AUC: area under the receiver operating characteristic curve.



A simplified point-based score was derived from the final model (Table 4). Age was assigned 1 point per 5-year increment above 60 years, sST2 was assigned 2 points per 5 ng/mL increment above 20 ng/mL, and eGFR was assigned 1 point per 10-unit decrease below 60 mL/min/1.73 m². For each component, values below the reference value for age and sST2, or above the reference value for eGFR, were assigned 0 points, and negative points were not allowed. In the derivation cohort, the observed total SAGE-HF score ranged from 0 to 12. Because no patient had a total score above 12 in this cohort, predicted probabilities were tabulated only across the observed score range. For future application, scores above 12 should be categorized as ≥12 and interpreted cautiously because risk estimates beyond the observed range would represent extrapolation and require external validation. The predicted probability of 30-day MACE increased progressively with higher total scores, ranging from 7.0% at a score of 0 to 60.1% at a score of 12 (Table 5). For clinical interpretability, provisional risk strata were defined as low risk (<10%), intermediate risk (10% - 30%), and high risk (>30%) according to predicted 30-day MACE probability. These strata corresponded to SAGE-HF scores of 0 to 1, 2 to 6, and 7 to 12, respectively. The observed MACE risk increased across these strata, from 2.2% in the low-risk group to 22.0% in the intermediate-risk group and 37.0% in the high-risk group. The corresponding exact binomial 95% CIs were 0.1% to 11.5%, 14.9% to 30.6%, and 24.3% to 51.3%, respectively (Table S5 in Multimedia Appendix 1). The simplified score also demonstrated acceptable discrimination and calibration for predicting 30-day MACE (Figure 3). Decision curve analysis suggested that the SAGE-HF score provided greater net benefit than the treat-none strategy and a more favorable net benefit than the treat-all strategy across much of the clinically relevant threshold-probability range, with the advantage over treat-all becoming more evident above approximately the 10% threshold (Figure S1 in Multimedia Appendix 1). In the exploratory benchmark analysis, the SAGE-HF simple score and the adjusted OPTIMIZE-HF in-hospital-derived score showed comparable performance for predicting 30-day postdischarge MACE. The SAGE-HF simple score had an apparent AUC of 0.727, bootstrap 95% CI 0.656 - 0.796, and bootstrap-corrected AUC of 0.726. The adjusted OPTIMIZE-HF-derived score had an apparent AUC of 0.731, bootstrap 95% CI of 0.659 - 0.805, and bootstrap-corrected AUC of 0.730. The difference in AUC between the 2 scores was minimal and not statistically significant (Tables S6 and S7 in Multimedia Appendix 1).

Table . Simplified point-based score derived from the final parsimonious model[a].

Predictor	Reference value	Increment	Assigned points
Age	60 years	+5 years	+1
sST2[b]	20 ng/mL	+5 ng/mL	+2
eGFR[c]	60 mL/min/1.73 m²	−10 mL/min/1.73 m²	+1

[a]Points were assigned by rounding the ratio of each clinically rescaled shrunken coefficient to the smallest rescaled coefficient.

[b]sST2: soluble suppression of tumorigenicity 2.

[c]eGFR: estimated glomerular filtration rate.

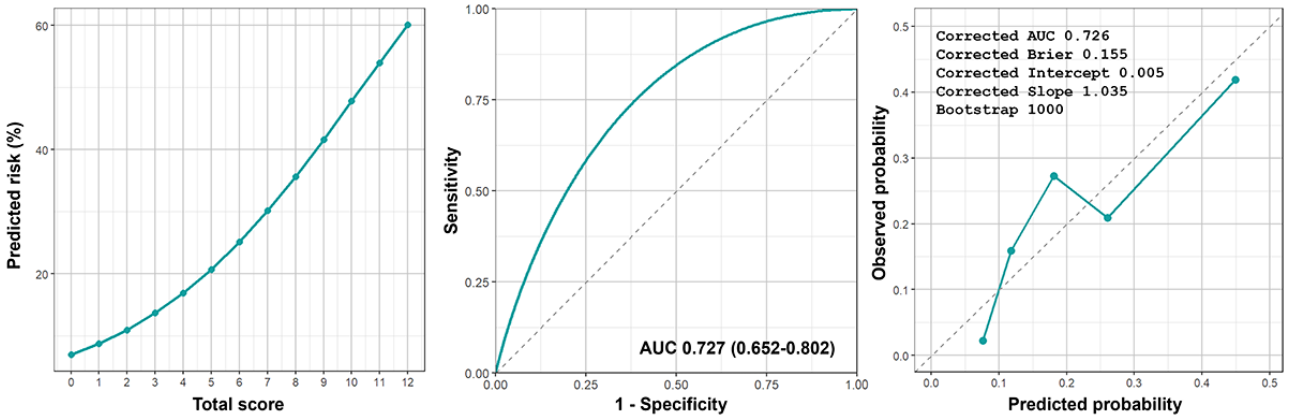
XSL•FO
RenderX

Table . Predicted probability of major adverse cardiovascular events (MACE) according to the total simplified risk score (SAGE-HF[a] score).

Total score	Predicted risk, % (95% CI)
0	7.0 (3.5 - 13.2)
1	8.8 (4.9 - 15.2)
2	11.0 (6.8 - 17.4)
3	13.7 (9.1 - 20.0)
4	16.9 (12.1 - 23.2)
5	20.7 (15.5 - 27.2)
6	25.2 (19.3 - 32.1)
7	30.2 (23.1 - 38.4)
8	35.7 (26.8 - 45.6)
9	41.6 (30.5 - 53.6)
10	47.8 (34.2 - 61.7)
11	54.0 (37.9 - 69.3)
12	60.1 (41.7 - 76.1)

[a]SAGE-HF: soluble ST2, age, and estimated glomerular filtration rate–heart failure.

Figure 3. Performance of the simplified risk score for predicting major adverse cardiovascular events (MACE). AUC: area under the receiver operating characteristic curve.



Discussion

Principal Findings

In this two-center prospective study, serum sST2 concentrations were significantly higher in patients with HF compared with non-HF controls. In the two-center prospective HF cohort, 21.6% (47/218) of the patients experienced 30-day postdischarge MACE, comprising 19.3% (42/218) HF rehospitalizations and 2.3% (5/218) deaths. Although hypertension was less frequent in the MACE group, this crude inverse association should be interpreted cautiously and should not be considered evidence of a protective effect. Hypertension was recorded as a binary history-based comorbidity and may not reflect hemodynamic status during the index hospitalization. This finding may reflect residual confounding, selection or collider bias, the high overall prevalence of hypertension in the cohort, and the small number of patients without hypertension. In a sensitivity analysis, admission systolic blood pressure was not associated with 30-day MACE (OR 0.99 per mm Hg, 95% CI 0.985 - 1.013; *P*=.89), supporting the interpretation that the inverse association for categorical hypertension was unlikely to represent a true protective effect. Patients with events were older and had higher admission sST2 and worse renal indices. Among prespecified candidate models, a parsimonious model including age, sST2, and eGFR achieved acceptable discrimination (AUC 0.741) and favorable internal validation with good calibration characteristics. A simplified integer score (SAGE-HF score) derived from the final model produced a clinically interpretable risk gradient, with predicted event probabilities rising from low single digits to approximately 60% across the observed score spectrum.

Age, renal function, and sST2 each contributed distinct prognostic information in this cohort. Patients with 30-day MACE were older than those without MACE (mean 74.9, SD 13.0 y vs 68.3, SD 12.1 y), and age remained independently associated with the end point, with an adjusted OR of 1.04 (95% CI 1.01 - 1.07) per year. This finding is consistent with the recognized vulnerability of older adults with HF, who often have greater comorbidity burden, frailty, and reduced physiologic reserve [26]. Renal dysfunction was also more

XSL•FO
RenderX

prominent in patients with events. The MACE group had lower eGFR than the non-MACE group (median 44.0, IQR 30.78-61.94 mL/min/1.73 m² vs 62.24, IQR 45.53-84.77 mL/min/1.73 m²), and eGFR remained independently associated with MACE, with an adjusted OR of 0.98 (95% CI 0.97 - 1.00) per 1 mL/min/1.73 m². This supports the clinical relevance of cardiorenal interaction in early postdischarge risk after HF hospitalization [27]. Admission sST2 was higher in patients who developed MACE than in those who did not (median 27.61, IQR 23.36-30.62 ng/mL vs 22.68, IQR 15.94-30.27 ng/mL) and remained independently associated with 30-day MACE after adjustment for age and eGFR, with an adjusted OR of 1.06 (95% CI 1.01 - 1.11) per ng/mL. This finding is consistent with prior literature linking higher sST2 concentrations to adverse outcomes in acute and chronic HF settings [9,28]. Mechanistically, sST2 reflects the activation of the IL-33/ST2 pathway and may capture myocardial stress, inflammation, and profibrotic remodeling [8,29]. In this cohort, the contribution of sST2 was observed despite the inclusion of renal function in the model, supporting its potential value as a complementary biomarker in short-term risk stratification.

A key finding of this study is that a 3-predictor model incorporating age, admission sST2, and eGFR provided clinically interpretable short-term risk stratification in a cohort with 47 MACE among 218 hospitalized patients with HF. The final model achieved an apparent AUC of 0.741 (95% CI 0.664 - 0.819), and its performance remained stable after internal validation with 1000 bootstrap resamples, with an optimism-corrected AUC of 0.725, Brier score of 0.155, calibration intercept of 0.009, and calibration slope of 0.930. These values indicate acceptable discrimination and calibration for an early postdischarge outcome, while also showing the expected degree of optimism in a relatively small derivation cohort. Compared with the 4-predictor model including serum sodium, the 3-predictor model had comparable or better optimism-corrected discrimination and a more favorable calibration slope, supporting the decision to prioritize parsimony.

The simplified SAGE-HF score further translated this model into a bedside format. In the derivation cohort, the predicted 30-day MACE probability increased from 7.0% at a score of 0 to 60.1% at a score of 12. When provisional risk strata were applied, the observed MACE risk was 2.2% in the low-risk group, 22.0% in the intermediate-risk group, and 37.0% in the high-risk group, with exact binomial 95% CIs of 0.1% to 11.5%, 14.9% to 30.6%, and 24.3% to 51.3%, respectively. This gradient is clinically relevant because the overall 30-day MACE rate was 21.6%, comprising 42 HF rehospitalizations and 5 deaths. However, because the low-risk stratum included only 1 event and its upper confidence limit exceeded 10%, the proposed <10% threshold should be interpreted as provisional rather than definitive.

These findings should be interpreted in the context of previous HF prediction literature. Prior studies have emphasized that 30-day readmission after HF hospitalization is difficult to predict because early events are driven by heterogeneous clinical and health-system factors [4]. Machine learning and traditional models for 30-day HF readmission have also shown variable performance and limited external validation across settings [30]. Against this background, the SAGE-HF model does not eliminate the need for clinical judgment, but it offers a low-burden and biologically interpretable approach using 3 readily available predictors. The inclusion of sST2 is consistent with literature showing that sST2 reflects myocardial stress, inflammation, fibrosis, and remodeling and has prognostic value in HF [28]. In this cohort, sST2 remained associated with MACE after adjustment for age and eGFR, supporting its potential role as a complementary biomarker rather than a replacement for standard clinical assessment.

From a postdischarge care perspective, the practical value of the SAGE-HF score lies in its ability to identify patients who may need more intensive early follow-up. This potential clinical usefulness was also supported by decision curve analysis, which showed a favorable net benefit of the SAGE-HF score across a clinically relevant range of threshold probabilities. Timely postdischarge care and structured transition-of-care strategies have been emphasized in HF management and have been associated with lower readmission rates in HF populations [31,32]. In this cohort, patients with scores ≥7 had an observed 30-day MACE risk of 37.0%, suggesting that this group may be considered for closer follow-up, medication optimization, early reassessment, or structured telephone monitoring after discharge. However, these risk categories remain exploratory and should not be used as definitive clinical thresholds until validated in larger and geographically diverse cohorts.

Limitations

This study has several strengths. The prospective design and inclusion of patients from 2 different centers enhance the reliability of the findings. The model was developed based on a prespecified and transparent strategy, with internal validation using bootstrap methods and comprehensive assessment of calibration. In addition, the inclusion of a non-HF comparison group allowed for an exploratory comparison of sST2 levels between patients with HF and individuals without HF. A Firth penalized logistic regression analysis was also performed to assess the robustness of the model and minimize potential bias related to the limited number of events. Furthermore, a simplified risk score was derived to facilitate clinical application, particularly in discharge planning and early postdischarge risk stratification. However, several limitations should be considered. First, the number of events was relatively small, which limited the number of predictors that could be included and may affect the stability of the model. To contextualize sample size adequacy for prediction-model development, an additional assessment using the pmsampsize framework was performed (Table S8 in Multimedia Appendix 1). With an anticipated C-statistic of 0.74, an outcome prevalence of 0.216, and 15 candidate predictor parameters before model reduction, the estimated minimum sample size was 1031 patients, corresponding to 223 events. In a sensitivity calculation using 19 candidate predictor parameters, the estimated requirement increased to 1306 patients, corresponding to 283 events. Because the present derivation cohort included 218 patients and forty-seven 30-day MACE, the study was underpowered for fitting a full multivariable prediction model containing all candidate predictors before reduction. Therefore, this study should be interpreted as an internally validated derivation study of a

XSL•FO
RenderX

deliberately parsimonious score rather than a fully powered development study for a large multivariable prediction model. Second, the follow-up period was limited to 30 days, which may not fully reflect longer-term outcomes after HF hospitalization. In addition, death before rehospitalization may act as a competing event when rehospitalization is analyzed separately. Because the dataset recorded the occurrence of 30-day MACE rather than complete component-specific time-to-event data, a Fine-Gray competing-risk analysis could not be performed appropriately. Third, short-term outcomes after discharge may also be influenced by health care system factors, such as access to follow-up, medication availability, and care coordination, which were not fully captured in this study. In addition, the counterintuitive inverse association between recorded hypertension and 30-day MACE may reflect residual confounding, selection or collider bias, and the limitation of using a history-based binary comorbidity variable rather than detailed longitudinal blood pressure, treatment, or hemodynamic data. Fourth, serum sST2 was measured using a research-use Elabscience ELISA kit rather than the US Food and Drug Administration–cleared Presage ST2 Assay. Therefore, absolute sST2 concentrations in this study should not be directly compared with Presage-based cohorts, and the Presage 35 ng/mL cutoff is not directly transferable. The observed associations and the SAGE-HF score should be interpreted as assay-specific findings that require external validation, ideally including assay harmonization or comparison with clinically validated sST2 platforms. Fifth, the non-HF comparison group was not formally matched to the HF cohort by age, sex, or BMI. Therefore, comparisons between patients with HF and non-HF comparison participants should be interpreted as exploratory and were not used in the development, validation, or performance assessment of the SAGE-HF score. Sixth, although this study included patients from 2 hospitals, the number of centers was small and the findings may not be generalizable to other regions, hospital levels, or HF populations with different etiology patterns. Center-level influence was assessed using a fixed-effect sensitivity analysis, but external validation in larger and geographically diverse multicenter cohorts remains necessary. In addition, formal comparison with some established HF risk scores was limited because several required predictors for these tools, particularly the OPTIMIZE-HF postdischarge mortality score, were not systematically collected in this cohort. Finally, the proposed risk thresholds are exploratory and were derived from the current cohort. They should not be used as definitive clinical decision thresholds before external validation. In addition, the low-risk stratum contained only 1 MACE. Although the observed event rate was low, the exact 95% CI extended above 10%, indicating that the proposed low-risk threshold requires confirmation in larger external validation cohorts. Further studies are needed to validate these findings in larger and more diverse populations, with longer follow-up and adequate event numbers. Such work may help clarify the role of this model in routine clinical practice.

Conclusions

In this two-center prospective cohort of patients hospitalized for HF, 21.6% (47/218) experienced 30-day postdischarge MACE. Admission age, sST2, and eGFR were independently associated with risk and were integrated into the SAGE-HF score, which demonstrated acceptable predictive performance. The score provided clear risk stratification with a graded increase in predicted MACE probability. External validation is warranted before routine clinical application.

Acknowledgments

We express our sincere gratitude to the Can Tho University of Medicine and Pharmacy, Thong Nhat Dong Nai General Hospital, and Can Tho Central General Hospital for creating favorable conditions for this study to be carried out. We confirm that no generative AI tools were used at any stage in the preparation, writing, or revision of this manuscript.

Funding

No financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: AVT, NKQL, VATP, TAKP
Formal analysis: NKQL, VATP, BTN
Methodology: AVT, NKQL, VATP, TAKP
Project administration: AVT, NKQL, VATP, TAKP
Supervision: AVT, TAKP
Validation: AVT, NKQL, VATP, TAKP, BTN, HTN, NHTT
Visualization: NKQL, BTN, HTN, NHTT
Writing – original draft: NKQL, BTN, HTN, NHTT
Writing – review and editing: AVT, TAKP

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figure presenting baseline comparisons, univariable predictors, prediction model performance, internal validation, risk stratification, benchmark comparison, sample size assessment, and decision curve analysis for the soluble ST2, age, and estimated glomerular filtration rate–heart failure score.
[DOCX File, 151 KB - cardio_v10i1e97879_app1.docx]

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
2. Doan QH, Tran AV, Dinh PTH, Nguyen LD. Enhancing quality of life and functional capacity in heart failure patients with iron deficiency: a regional cross-sectional study on intravenous iron therapy in Vietnam. Medicine (Baltimore) 2025 Aug 22;104(34):e44060. [doi: 10.1097/MD.0000000000044060] [Medline: 40859504]
3. Greene SJ, Fonarow GC, Vaduganathan M, Khan SS, Butler J, Gheorghiade M. The vulnerable phase after hospitalization for heart failure. Nat Rev Cardiol 2015 Apr;12(4):220-229. [doi: 10.1038/nrcardio.2015.14] [Medline: 25666406]
4. Dharmarajan K, Hsieh AF, Lin Z, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA 2013 Jan 23;309(4):355-363. [doi: 10.1001/jama.2012.216476] [Medline: 23340637]
5. Bueno H, Ross JS, Wang Y, et al. Trends in length of stay and short-term outcomes among Medicare patients hospitalized for heart failure, 1993-2006. JAMA 2010 Jun 2;303(21):2141-2147. [doi: 10.1001/jama.2010.748] [Medline: 20516414]
6. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. JAMA Cardiol 2017 Feb 1;2(2):204-209. [doi: 10.1001/jamacardio.2016.3956] [Medline: 27784047]
7. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022 May 3;145(18):e895-e1032. [doi: 10.1161/CIR.0000000000001063] [Medline: 35363499]
8. Kakkar R, Lee RT. The IL-33/ST2 pathway: therapeutic target and novel biomarker. Nat Rev Drug Discov 2008 Oct;7(10):827-840. [doi: 10.1038/nrd2660] [Medline: 18827826]
9. Januzzi JL Jr, Peacock WF, Maisel AS, et al. Measurement of the interleukin family member ST2 in patients with acute dyspnea: results from the PRIDE (Pro-Brain Natriuretic Peptide Investigation of Dyspnea in the Emergency Department) study. J Am Coll Cardiol 2007 Aug 14;50(7):607-613. [doi: 10.1016/j.jacc.2007.05.014] [Medline: 17692745]
10. Sciatti E, Merlo A, Scangiuzzi C, et al. Prognostic value of sST2 in heart failure. J Clin Med 2023 Jun 11;12(12):3970. [doi: 10.3390/jcm12123970] [Medline: 37373664]
11. Riccardi M, Myhre PL, Zelniker TA, Metra M, Januzzi JL, Inciardi RM. Soluble ST2 in heart failure: a clinical role beyond B-type natriuretic peptide. J Cardiovasc Dev Dis 2023 Nov 17;10(11):468. [doi: 10.3390/jcdd10110468] [Medline: 37998526]
12. Matyar S, Açıkalın Akpınar A, Dişel NR, et al. Prognostic value of sst2 in long-term mortality in acute heart failure. Acta Cardiol 2024 Oct;79(8):924-934. [doi: 10.1080/00015385.2024.2406683] [Medline: 39317343]
13. Bosco E, Hsueh L, McConeghy KW, Gravenstein S, Saade E. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. BMC Med Res Methodol 2021 Nov 6;21(1):241. [doi: 10.1186/s12874-021-01440-5] [Medline: 34742250]
14. Yoo JE, Jeong SM, Yeo Y, et al. Smoking cessation reduces the risk of heart failure: a nationwide cohort study. JACC Heart Fail 2023 Mar;11(3):277-287. [doi: 10.1016/j.jchf.2022.07.006] [Medline: 36647926]
15. Wong B, Moore A, McDonald K, Ledwidge M. Alcohol consumption and progression of heart failure in those at risk for or with pre-heart failure. J Card Fail 2024 Dec;30(12):1555-1563. [doi: 10.1016/j.cardfail.2024.04.003] [Medline: 38679412]
16. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015 Jan;28(1):1-39. [doi: 10.1016/j.echo.2014.10.003] [Medline: 25559473]
17. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 5;150(9):604-612. [doi: 10.7326/0003-4819-150-9-200905050-00006] [Medline: 19414839]
18. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138. [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med 2019 Dec 16;17(1):230. [doi: 10.1186/s12916-019-1466-7] [Medline: 31842878]

XSL•FO
RenderX

20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]
21. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
22. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II—binary and time-to-event outcomes. Stat Med 2019 Mar 30;38(7):1276-1296. [doi: 10.1002/sim.7992] [Medline: 30357870]
23. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80(1):27-38. [doi: 10.1093/biomet/80.1.27]
24. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002 Aug 30;21(16):2409-2419. [doi: 10.1002/sim.1047] [Medline: 12210625]
25. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26(6):565-574. [doi: 10.1177/0272989X06295361] [Medline: 17099194]
26. Díez-Villanueva P, Jiménez-Méndez C, Alfonso F. Heart failure in the elderly. J Geriatr Cardiol 2021 Mar 28;18(3):219-232. [doi: 10.11909/j.issn.1671-5411.2021.03.009] [Medline: 33907552]
27. Mullens W, Martens P, Testani JM, et al. Renal effects of guideline-directed medical therapies in heart failure: a consensus document from the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2022 Apr;24(4):603-619. [doi: 10.1002/ejhf.2471] [Medline: 35239201]
28. Aimo A, Januzzi JL Jr, Bayes-Genis A, et al. Clinical and prognostic significance of sST2 in heart failure: JACC review topic of the week. J Am Coll Cardiol 2019 Oct 29;74(17):2193-2203. [doi: 10.1016/j.jacc.2019.08.1039] [Medline: 31648713]
29. Feng Y, He LQ. Soluble ST2: a novel biomarker for diagnosis and prognosis of cardiovascular disease. Curr Med Sci 2024 Aug;44(4):669-679. [doi: 10.1007/s11596-024-2907-x] [Medline: 39096477]
30. Yu MY, Son YJ. Machine learning-based 30-day readmission prediction models for patients with heart failure: a systematic review. Eur J Cardiovasc Nurs 2024 Oct 21;23(7):711-719. [doi: 10.1093/eurjcn/zvae031] [Medline: 38421187]
31. Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA 2010 May 5;303(17):1716-1722. [doi: 10.1001/jama.2010.533] [Medline: 20442387]
32. Salah HM, Ambrosy AP, Biegus J, et al. Optimizing pre-to-post discharge transition of care in patients hospitalized for heart failure—part 3 of the international expert opinion series on acute heart failure management. J Card Fail 2025 Sep 28:S1071-9164(25)00433-6. [doi: 10.1016/j.cardfail.2025.09.020] [Medline: 41027508]

Abbreviations

ADHF: acute decompensated heart failure
AIC: Akaike information criterion
AUC: area under the receiver operating characteristic curve
eGFR: estimated glomerular filtration rate
ELISA: enzyme-linked immunosorbent assay
HF: heart failure
HFmrEF: heart failure with mildly reduced ejection fraction
HFpEF: heart failure with preserved ejection fraction
HFrEF: heart failure with reduced ejection fraction
IL: interleukin
LVEF: left ventricular ejection fraction
MACE: major adverse cardiovascular events
OR: odds ratio
SAGE-HF: soluble ST2, age, and estimated glomerular filtration rate–heart failure
sST2: soluble suppression of tumorigenicity 2
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence

XSL•FO
RenderX

Edited by A Coristine; submitted 11.Apr.2026; peer-reviewed by D Nikolić, H Lv; revised version received 18.Jun.2026; accepted 19.Jun.2026; published 29.Jul.2026.

Please cite as:
Tran AV, La NKQ, Phan VAT, Nguyen BT, Nguyen HT, Thai NHT, Pham TAK
Development and Internal Validation of the Soluble ST2, Age, and Estimated Glomerular Filtration Rate–Heart Failure Score for Predicting 30-Day Major Adverse Cardiovascular Events After Heart Failure Hospitalization in a Two-Center Vietnamese Cohort: Prospective Cohort Study
JMIR Cardio 2026;10:e97879
URL: https://cardio.jmir.org/2026/1/e97879
doi:10.2196/97879

© An Viet Tran, Nguyen Khoi Quoc La, Van Anh Thi Phan, Bao The Nguyen, Hoa Thai Nguyen, Nhung Hong Thi Thai, Tho Anh Kieu Pham. Originally published in JMIR Cardio (https://cardio.jmir.org), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility of Mobile Single-Lead Electrocardiogram Screening for Atrial Fibrillation: Multicenter Cross-Sectional Study

Robert Fenske[1], MD; Denise Rossa[1], MD; Christian Butter[1], Prof Dr med; Felix Muehlensiepen[2], PhD; Jonathan Nübel[1], MD; Daniel Meretz[1], MD; Oliver Ritter[3], Prof Dr med; Sebastian Spethmann[4,5], Prof Dr med; Rene Mantke[6], Prof Dr med; Mario Zerbaum[7], MD; Nico Draheim[8], MD; Nicole Lindenberg-Krause[9], MD; Anja Haase-Fielitz[1,2,10], Prof Dr rer medic

[1]Heart Center Brandenburg, Brandenburg Medical School (MHB) Theodor Fontane, Ladeburger Str. 17, Bernau, Brandenburg, Germany

[10]Institute of Social Medicine and Health System Research, Otto von Guericke University Magdeburg, Madgeburg, Germany

[2]Center for Health Services Research (ZVF-BB), Faculty of Health Sciences Brandenburg, Brandenburg Medical School (MHB) Theodor Fontane, Rüdersdorf, Brandenburg, Germany

[3]Department of Cardiology, Nephrology and Pneumology, Faculty of Health Sciences Bradenburg, Brandenburg Medical School Theodor Fontane, University Hospital Brandenburg, Brandenburg an der Havel, Germany

[4]Department of Cardiology, Angiology and Intensive Care Medicine, Campus Mitte, German Heart Center at Charité – University Medical Center Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[5]DZHK (German Centre for cardiovascular research), Partner site, Berlin, Germany

[6]Department of Surgery, University Hospital Brandenburg, Brandenburg Medical School Theodor Fontane, Brandenburg, Germany

[7]General practitioner Dr. Zerbaum & colleagues MVZ, Petersilienstraße 12, Brandenburg, Germany

[8]Family practice Nico Draheim, Fehrbelliner Str. 18, KST 19630044, Wichmannstrasse 18, Neuruppin, Germany

[9]General medical practice Dr. Lindenberg, Berliner Allee 18a, Werneuchen, Germany

Corresponding Author:
Anja Haase-Fielitz, Prof Dr rer medic
Heart Center Brandenburg, Brandenburg Medical School (MHB) Theodor Fontane, Ladeburger Str. 17, Bernau, Brandenburg, Germany

Abstract

Background: Atrial fibrillation (AF) is a major contributor of thromboembolic stroke. However, early detection remains challenging and is severely hindered by structural limitations in health care infrastructure. While current clinical guidelines recognize a 30-second single-lead electrocardiogram (ECG) as a standard diagnostic tool, the actual deployment and real-world utility of mobile screening technologies across health care settings remain poorly understood.

Objective: The primary objective of this study was to evaluate the real-world feasibility of a mobile single-lead ECG for the detection of AF in routine inpatient and outpatient settings. Feasibility was evaluated by the proportion of successfully completed recordings, patient acceptance, and the occurrence and characteristics of technical challenges during deployment.

Methods: This multicenter, cross-sectional pragmatic implementation study enrolled patients across 3 university hospitals and 3 primary care practices. Screening was performed using an US Food and Drug Administration (FDA)–approved single-lead ECG device (AliveCor KardiaMobile). Patients with an automated device reading indicating "Possible AF" and no prior diagnosis of AF underwent a confirmatory 12-lead ECG. Patient awareness and acceptance were evaluated using an exploratory, descriptive questionnaire.

Results: A total of 237 patients were enrolled (194 inpatients and 43 outpatients). The device algorithm provided an automatic interpretation in 95.3% (41/43) of the outpatients and 85.6% (166/194) of inpatients, resulting in an overall automated detection of AF in 21.5% (51/237) of the cohort. The algorithm identified newly suspected AF in 1.7% (4/237) of patients without a prior AF diagnosis. Of these, 1 case underwent same-day confirmatory 12-lead ECG and was confirmed as AF. Technical challenges occurred in 16.0% (38/237) of all recordings, with a numerical higher rate observed in the inpatient setting (34/194, 17.5% vs 4/43, 9.3% in outpatients; *P*=.25). Only 6.3% (15/237) of patients reported regular use of telemedicine applications.

Conclusions: Mobile single-lead ECG screening is feasible in routine inpatient and outpatient settings and identified a substantial prevalence of AF, which increased with age. The high detection rate underscores the potential of wearable ECG devices to facilitate early detection of AF. However, the absence of systematic confirmatory ECG testing and limited patient engagement with telemedicine applications highlight challenges in real-world implementation.

(JMIR Cardio 2026;10:e92799) doi:10.2196/92799

XSL•FO
RenderX

KEYWORDS

atrial fibrillation; screening; mobile health; single-lead ECG; AliveCor; telemedicine; electrocardiogram

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia and is associated with an increased risk of stroke, cognitive decline, and reduced quality of life. The global prevalence of AF has increased substantially over the past 3 decades, with an estimated 60 million cases worldwide [1]. Individuals with AF are more likely to be hospitalized for both cardiovascular and noncardiovascular causes [2]. With the aging population in Germany and globally, the burden of AF is expected to grow further. In Germany, the number of hospitals has decreased by 15% and hospital beds by 30% since the early 1990s [3]. In rural and underserved regions, access to health care remains limited, with travel times to hospitals or outpatient clinics exceeding 45 minutes [4]. Furthermore, the geographic density of specialized outpatient cardiologists in Brandenburg is significantly below the German national average, placing a disproportionate burden on primary care practitioners to manage complex chronic conditions such as AF [5].

Current cardiology guidelines recommend either a standard 12-lead electrocardiogram (ECG) or a single-lead ECG recording at least 30 seconds for the diagnosis of AF. In addition, opportunistic screening is also advised for patients aged 65 years and over [6]. Digital technologies such as telemedicine are transforming health care delivery by improving cross-sectoral care [7]. The increasing use of smartphones and wearables also offers a new, practical, and cost-effective way to improve AF screening and increase detection rates. These tools are accessible at any time, can be operated without prior medical knowledge, and are well-suited for documenting sporadic, short-term arrhythmias [8]. A recent study conducted in Canadian primary care settings demonstrated that opportunistic AF screening results in a significant reduction in downstream adverse health outcomes and future health expenditures [9]. However, while most validation studies of mobile single-lead ECG algorithms have been conducted under tightly controlled idealized conditions, their performance in real-world, pragmatic clinical settings remains poorly understood. There is a critical need to evaluate how these tools function in routine, cross-sectoral practice, where patient populations are heterogeneous, workflows are complex, and resources are limited.

The primary aim of this study was to assess the feasibility and real-world implementation of opportunistic AF screening using a mobile single-lead ECG device in diverse inpatient and outpatient care settings.

Methods

Study Design

This multicenter, cross-sectional study was conducted across 3 inpatient departments at the Brandenburg Medical School (Bernau, Brandenburg and Neuruppin) and 3 primary care practices. Patients were screened for AF using a mobile single-lead ECG (AliveCor KardiaMobile). The study was designed and implemented as part of a student-led research project within the Science Internship Program at the Brandenburg Medical School.

Study Population

Eligible patients were adults aged 65 years or older, recruited on September 29, 2021, World Heart Day. Patients who did not provide written informed consent or were under 65 years of age were excluded from the study. No additional exclusion criteria were applied. Inpatients were recruited from the departments of cardiology, surgery, internal medicine and oncology at 3 sites: Bernau, Brandenburg and Neuruppin. Outpatients were recruited from 3 primary care practices of general practitioners at the same geographic locations.

Primary and Secondary Outcomes

The primary outcome was to evaluate the feasibility and real-world implementation of opportunistic AF screening using a mobile single-lead ECG device, assessed by the proportion of successfully completed automated recordings.

Feasibility of the mobile single-lead ECG screening was assessed using 3 key indicators: the proportion of successfully completed recordings, patient acceptance, and the occurrence and characteristics of technical challenges during deployment. A successful recording was defined as a 30-second ECG with a clear, interpretable trace and no signal artifacts. Patient acceptance was evaluated through self-reported feedback collected via a structured questionnaire (Multimedia Appendix 1), including willingness to repeat the screening and perceived ease of use. Technical challenges including signal artifacts, unclassifiable tracing, and device-related issues were documented and categorized by type (eg, conductivity problems, patient movement, technical errors).

Secondary outcomes comprised an exploratory assessment of concordance between algorithm classifications and chart-documented AF history. Because a contemporaneous reference-standard ECG was not systematically performed in all participants, no formal diagnostic accuracy analysis was undertaken.

Study Intervention and Data Collection

Screening was conducted using a commercially available, US Food and Drug Administration (FDA)–approved handheld single-lead ECG device (AliveCor KardiaMobile, version 5.35.1). Patients placed the index fingers of both hands onto the integrated electrodes for a 30-second recording. The mobile single-lead ECG was used as the primary screening tool for all patients. Patients who received an automated alert of “Possible AF” and had no prior diagnosis of AF underwent a confirmatory 12-lead ECG. A standard 12-lead ECG was not performed systematically across the entire cohort. Subsequent clinical management, including physician consultations and the initiation of oral anticoagulation therapy, were managed directly by the

XSL•FO
RenderX

attending clinical staff on the same day, in accordance with standard care pathways.

No repeat ECGs were systematically performed. When a recording was flagged as "Unclassified" or "Unreadable," clinical staff performed basic, immediate troubleshooting, including wiping the device sensors, repositioning the patient's fingers, and asking the patient to relax or minimize movement. If these steps did not resolve the issue within a short time, the recording was accepted as inconclusive and not repeated.

Health literacy and digital readiness were assessed using a self-designed 12-item exploratory questionnaire. The instrument was conceptually informed by established frameworks, including the European Health Literacy Survey and the Digital Health Literacy Instrument. The questionnaire combined closed-ended items (multiple-choice and multicategorical checkboxes) to capture baseline demographics, pre-existing cardiac history, and potential physical or psychological barriers to technology adoption. Open-ended, free-text responses were intentionally included to explore patient awareness of AF symptoms, enabling qualitative insights into symptom recognition and health-related perceptions.

Statistical Analysis

Normality of continuous variables was assessed graphically using histograms. Normally distributed data were compared using the *t* test or, in cases of heterogeneous variances, the Welch test. Comparisons of nonnormally distributed data were conducted using the Mann-Whitney *U* test or the Wilcoxon test (for paired data). Categorical variables were compared using the chi-square test. Exploratory performance metrics of the automated algorithm including sensitivity and specificity were calculated by cross-referencing algorithm outputs with established, chart-documented diagnoses of AF. A *P* value of <.05 was considered statistically significant. All data were analyzed using SPSS version 29.0 (IBM Corp).

Ethical Considerations

The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by the Institutional Review Board of the Brandenburg Medical School (E-01-2020122). Informed consent was obtained prior to study participation from all individuals. Patients received detailed information regarding the objectives and procedures of the study, data handling, and storage practices, as well as potential health risks associated with or following the examinations. By consenting, they approved the collection, processing, and analysis of their personal data. At any point, and without the need to provide justification, participants retained the right to withdraw from the study, which would entail the erasure of all personal data and its copies. No financial or material incentives were offered for participation.

Results

Overview

On World Heart Day 2021, all patients hospitalized at the departments of cardiology, surgery, internal medicine and oncology, as well as all outpatients with scheduled appointments at 3 primary care practices, were invited to participate in the study. A total of 43 outpatients and 194 inpatients were enrolled. Using the KardiaMobile ECG, AF was detected in 13.9% (6/43) of outpatients and in 23.2% (45/194) of inpatients, resulting in an overall AF detection rate of 21.5% (51/237). Among these, 4 (1.7%) patients without a prior history of AF received the automated device interpretation indication "Possible AF" and were therefore considered newly suspected AF cases. All 4 patients were female. One of these cases underwent same-day 12-lead ECG confirmation, and anticoagulation therapy was initiated accordingly.

Patients

Demographic characteristics, comorbidities, and medication use among inpatients and outpatients are shown in Table 1. The overall median age was 73 (IQR 66 - 82) years and 48% (114/237) of patients were female. Inpatients were older and more frequently presented with heart valve disease and heart failure compared to outpatients (Table 1). The prevalence of known AF (91 inpatients, 6 outpatients) increased with age in both groups: from 18.7% (17/91) (65 - 69 years) to 45.0% (41/91) (80 - 89 years) among inpatients.

XSL•FO
RenderX

Table . Demographic data and comorbidities in outpatients and inpatients.

Variable	Outpatients (n=43)	Inpatients (n=194)	*P* value
Demographics, mean (IQR)			
Age (years)	70.0 (65.0-76.0)	75.0 (68.0-82.0)	.26
Heart rate (beats/min)	78.0 (70.0-85.0)	80.0 (70.0-89.0)	.81
CHA_2DS_2-VASc[a] score (points)	3.0 (2.0-4.0)	4.0 (3.0-5.0)	.10
Comorbidities, n (%)			
Arterial hypertension	32 (74.4)	144 (74.2)	.98
Coronary heart disease	15 (34.9)	76 (39.2)	.60
Diabetes mellitus	17 (39.5)	63 (32.5)	.38
Heart failure	6 (14.0)	72 (37.1)	.003
Pre-existing atrial fibrillation	6 (13.6)	87 (45.1)	<.001
Prior myocardial infarction	4 (9.3)	18 (9.3)	>.99
Heart valve disease	5 (11.6)	89 (45.9)	<.001
Chronic kidney disease	8 (18.6)	49 (25.3)	.36
Peripheral vascular disease	15 (34.9)	94 (48.5)	.11
Asthma bronchial	3 (7.0)	11 (5.7)	.74
Chronic obstructive pulmonary disease	7 (16.3)	21 (10.8)	.32
Medication, n (%)			
Beta-blockers	17 (37.8)	129 (66.5)	<.001
AT-1 receptor antagonists	14 (32.6)	63 (32.5)	.99
ACE[b] inhibitors	12 (27.9)	60 (30.9)	.70
Aldosterone antagonists	3 (7.0)	39 (20.1)	.04
Calcium channel blockers	7 (16.3)	49 (25.3)	.21
Diuretics	10 (23.3)	104 (53.6)	<.001
Neprilysin inhibitors (ARNI)[c]	0 (0)	12 (6.2)	.09
Statins	19 (44.2)	97 (50)	.49
Acetylsalicylic acid	13 (30.2)	57 (29.4)	.91
DOACs[d]	5 (11.6)	66 (34.0)	.004
Thyroid medication	7 (16.3)	26 (13.4)	.62

[a]CHA_2DS_2-VASc: congestive heart failure, hypertension, age, diabetes mellitus, prior stroke or TIA or thromboembolism, vascular disease, age, sex category.

[b]ACE: angiotensin-converting enzyme.

[c]ARNI: angiotensin receptor/neprilysin inhibitor.

[d]DOAC: direct oral anticoagulant.

Patients with AF were more likely to have additional cardiovascular comorbidities, including prior myocardial infarction, coronary heart disease, or heart failure, compared to patients without AF (Table 2). Patients with AF also received direct oral anticoagulants more frequently (particularly apixaban and rivaroxaban), as well as beta-blockers, angiotensin-converting enzyme inhibitors, aldosterone antagonists, and diuretics, compared to patients without AF.

XSL•FO
RenderX

Table . Demographic data and comorbidities in patients with and without atrial fibrillation detection using KardiaMobile.

Variables	Patients with sinus rhythm (n=147)	Patients with AF detection (n=51)	*P* value
Demographics			
Age (years), median (IQR)	72.0 (66.0-78.0)	78.0 (75.0-83.0)	.004
Age >65 years, n (%)	120 (81.6)	47 (92.2)	.04
Age >75 years, n (%)	50 (34.0)	38 (74.5)	<.001
Gender (female), n (%)	85 (57.8)	28 (54.9)	.82
Heart rate (beats/min)	76.0 (69.0-85.0)	87.0 (73.0-95.0)	.03
CHA_2DS_2-VASc[a] score (points)	4.0 (2.0-5.0)	5.0 (4.0-5.0)	.002
Comorbidities, n (%)			
Arterial hypertension	101 (68.7)	42 (82.4)	.04
Coronary heart disease	48 (32.7)	23 (45.1)	.09
Diabetes mellitus	46 (31.3)	16 (31.4)	.93
Heart failure	34 (23.4)	28 (54.9)	<.001
Prior myocardial infarction	14 (9.5)	3 (6.0)	.44
Heart valve disease	43 (29.3)	32 (62.7)	<.001
Chronic kidney disease	29 (19.7)	16 (31.3)	.07
Peripheral vascular disease	61 (41.5)	25 (49.0)	.06
Asthma bronchial	7 (4.8)	3 (6.0)	.73
Chronic obstructive pulmonary disease	15 (10.2)	6 (11.8)	.72
Medication, n (%)			
Beta-blockers	81 (55.1)	43 (84.3)	<.001
AT-1 receptor antagonists	50 (34.0)	13 (25.5)	.29
ACE[b] inhibitor	46 (31.3)	17 (33.3)	.72
Aldosterone antagonists	16 (10.9)	20 (39.2)	<.001
Calcium channel blockers	42 (28.6)	4 (7.8)	.003
Diuretics	58 (39.5)	33 (64.7)	.001
Neprilysin inhibitor	3 (2.0)	6 (11.8)	.04
Statins	72 (49.0)	23 (45.1)	.72
Acetylsalicylic acid	53 (36.1)	6 (11.8)	.001
DOACs[c]	25 (17.0)	37 (72.5)	<.001
Thyroid medication	18 (12.2)	6 (11.8)	.96
Laboratory parameters, median (IQR)			
Potassium (mmol/l)	4.28 (3.98-4.52)	4.19 (3.81-4.43)	.66
Sodium (mmol/l)	139.0 (137.0-141.0)	138.5 (136.0-140.6)	.79
Creatinine (µmol/l)	91.2 (64.0-106.0)	98.9 (75.0-111.0)	.39
Thyroid-stimulating hormone (µlU/ml)	2.1 (0.9-2.4)	1.7 (1.1-1.9)	.48

[a]CHA_2DS_2-VASc: congestive heart failure, hypertension, age, diabetes mellitus, prior stroke or TIA or thromboembolism, vascular disease, age, sex category.

[b]ACE: angiotensin-converting enzyme.

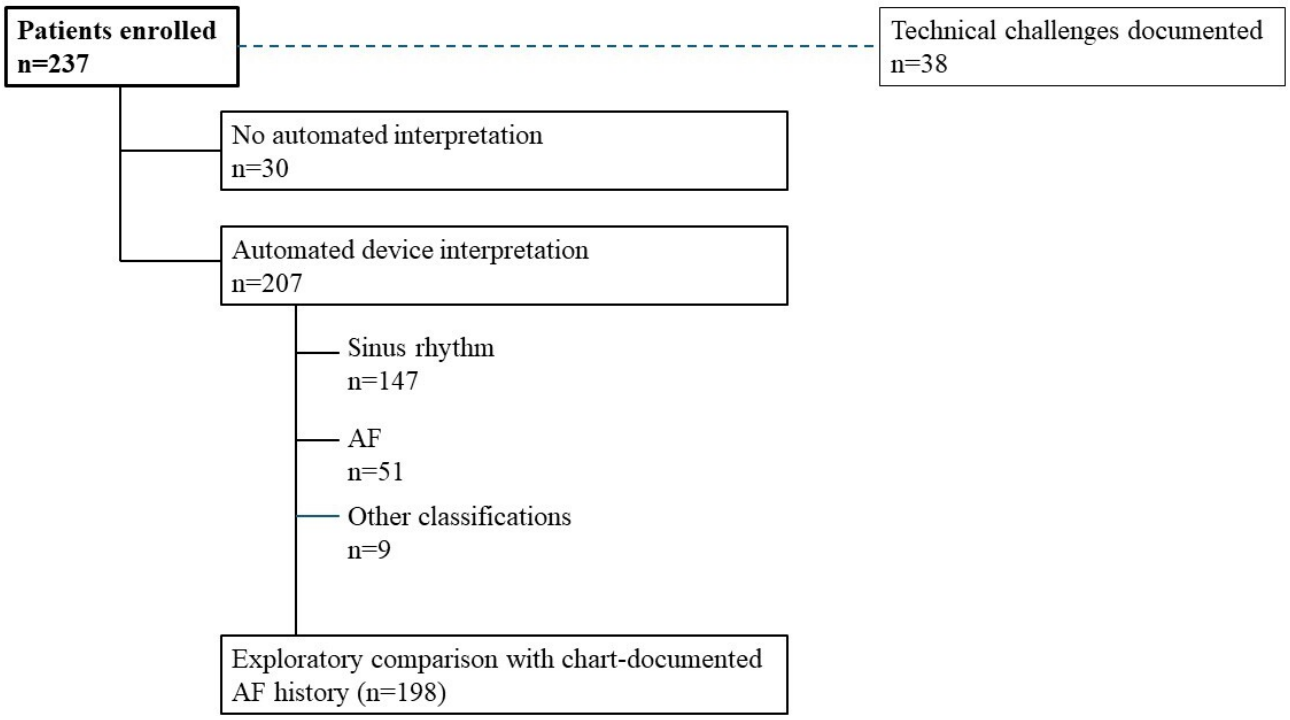
[c]DOAC: direct oral anticoagulant.

XSL•FO
RenderX

Feasibility of the Mobile Single-Lead ECG Screening

The device algorithm provided an automated interpretation in 85.6% (166/194) of inpatient recordings and 95.3% (41/43) of outpatient recordings, demonstrating successful deployment in most patients (Figure 1). Technical challenges were documented in 38 (16.0%) recordings. These challenges did not necessarily prevent generation of automated device interpretation. The rate of technical challenges was slightly higher in the inpatient sector (34/194, 17.5%) compared to outpatients (4/43, 9.3%; *P*=.25). In 5 patients (2.1% of cases), no specific issue was documented. Technical challenges were primarily related to patient-related factors such as movement, poor electrode contact, or comorbidities and device-related issues including connectivity problems and software glitches. Among inpatients, the most frequent problems were related to comorbidities or compliance (11/194, 5.7%), technical errors (11/194, 5.7%), and conductivity issues (9/194, 4.6%). In outpatients, conductivity problems (4/43, 9.3%) were the most common reported issues.

Figure 1. Flow of study participants and analysis populations. Overview of participant inclusion, automated device interpretation results, documented technical challenges, and the subset included in the exploratory comparison with chart-documented atrial fibrillation (AF) history.



Exploratory Comparison With Chart-Documented AF

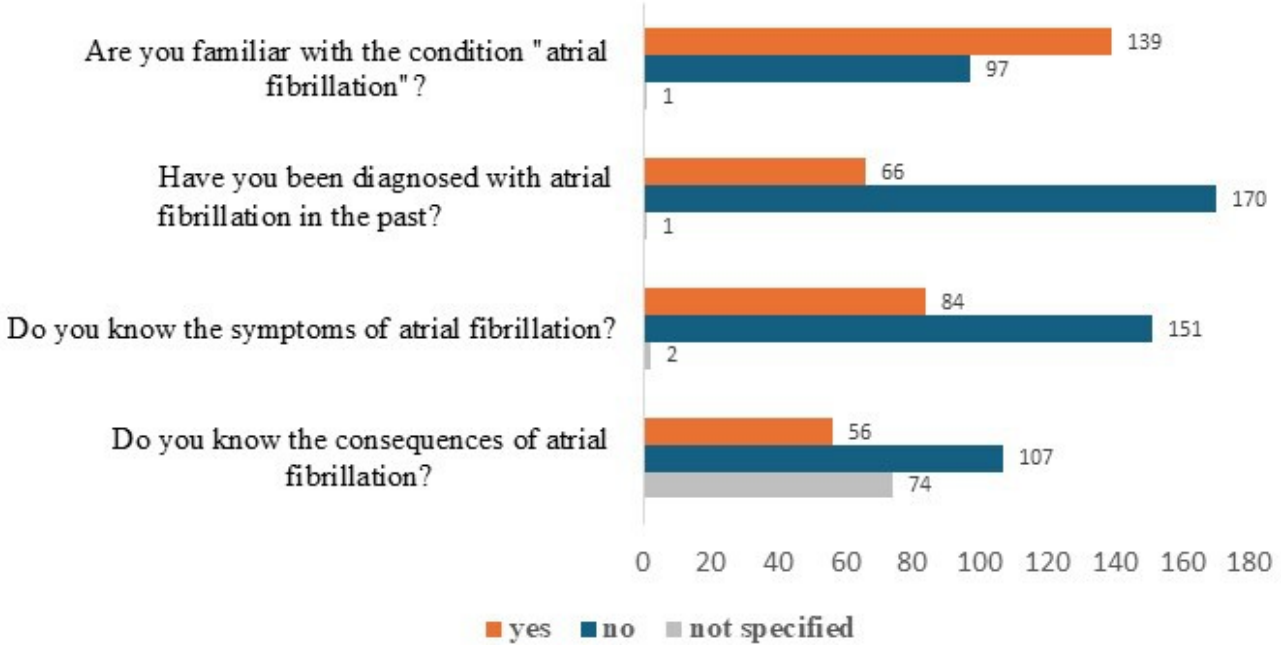
Of the 207 recordings that yielded an automated device interpretation, 198 were classified as either sinus rhythm (n=147) or AF (n=51) and were therefore included in the exploratory comparison with chart-documented AF history. The remaining recordings received alternative automated classifications and were excluded from this analysis.

Use of Telemedicine and Patient Knowledge Regarding AF

In this study, 4.6% (9/194) of inpatients and 14.0% (6/43) of outpatients reported that they regularly use telemedicine applications such as health apps, telephone consultations, smartwatches, or contactless blood glucose monitoring. Patient acceptance of home-based mobile ECG screening was generally high; 56.5% (134/237) of participants expressed interest in using a 30-second ECG at home, and only 42.6% (101/237) declined. The most commonly reported reasons for refusal were visual impairment (29/237, 12.2%), skepticism toward digital health technologies (24/237, 10.1%), lack of interest (20/237, 8.4%), and concerns about technical difficulties (15/237, 6.3%). Patients' knowledge about AF is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Patient knowledge about atrial fibrillation (absolute numbers).



Furthermore, among inpatients with AF documented in their medical records, 42.5% (37/87) were aware of their diagnosis. Of those, 89.2% (33/37) reported experiencing symptoms within the past 6 months.

Among outpatients, 6 of 43 (14.0%) reported a diagnosis of AF. The condition “atrial fibrillation” was recognized by 46.5% (20/43) of outpatients. Of these, 30.2% (13/43) were familiar with the symptoms, and 18.6% (8/43) were aware of potential complications, including stroke and myocardial infarction.

Discussion

Principal Findings

This multicenter study evaluated the feasibility and real-world performance of opportunistic AF screening using a mobile single-lead ECG (KardiaMobile) in routine inpatient and outpatient settings. Our findings demonstrate that mobile ECG screening is feasible in diverse clinical environments, with an overall AF detection rate of 21.5% (51/237), particularly among older adults. However, the study also reveals significant challenges in technical performance and patient-centered implementation, underscoring the gap between idealized validation and real-world application.

The observed prevalence of 21.5% substantially exceeds the estimated 1% - 2% prevalence of AF in the general German population and aligns with findings from high-risk cohorts. The elevated rate of AF is likely attributable to selection bias, as our cohort included hospitalized patients and patients visiting a general practitioner, groups inherently at higher risk for cardiovascular disease. The prevalence of AF increased with age among inpatients, reinforcing the value of age-targeted screening in older populations

In previous opportunistic screening studies using single-lead ECG, the incidence of newly diagnosed AF ranged from 0.49% [10] to 6.5% [11], depending on the baseline clinical risk profiles of the population. Our newly diagnosed rate of 1.7% is consistent with large-scale community screening initiatives, such as the study by Lowres et al [12], which reported a detection rate of 1.5% using a smartphone-based interface. However, in our study, only 1 case underwent confirmatory 12-lead ECG on the same day, reflecting both the pragmatic design of the study and the absence of systematic confirmatory testing for all screen-positive participants. Therefore, this figure should be interpreted as the rate of screen-detected suspected AF rather than confirmed newly diagnosed AF.

As expected, patients with AF had significantly higher median CHA_2DS_2-VASc (congestive heart failure, hypertension, age, diabetes mellitus, prior stroke or TIA or thromboembolism, vascular disease, age, sex category) scores than those in sinus rhythm. However, unlike the study by Lowres et al [12], we found no significant difference in risk scores between newly suspected and known AF cases. Notably, all our newly diagnosed patients were female, a finding that contradicts broader epidemiological data and is likely a statistical artifact due to the small sample size and the study’s recruitment focus on high-risk settings.

The growing prevalence of smartphone—already used by 65% of adults over 70 in Germany—provides a strong foundation for scalable digital screening [13]. The eBRAVE-AF study demonstrated that smartphone-based screening can more than double AF detection rates in individuals aged 50 years and older compared to routine care [14]. Similarly, the 12-month REHEARSE-AF study reported high patient adherence, with 90% of participants performing at least 1 ECG per week [15]. The STROKESTOP study, a multicenter, unmasked randomized controlled trial involving over 13,000 participants aged 75 and 76 years in Sweden, demonstrated that intermittent ECG screening could quadruple the detection rate of AF compared to single-time-point assessments [16]. STROKESTOP emphasized that high patient adherence and correct device use are critical for successful screening. In contrast to

XSL•FO
RenderX

STROKESTOP, which required 14-day continuous monitoring, our study relied on a single 30-second recording. This difference underscores the trade-off between diagnostic sensitivity and operational feasibility. While longer monitoring increases detection rates, brief, opportunistic screening may offer a more scalable and sustainable model for implementation in diverse health care environments. While studies like STROKESTOP [16] and eBRAVE-AF [14] evaluate the efficacy of structured screening programs under controlled conditions, our study focused on the real-world feasibility and implementation of mobile ECG screening within routine clinical workflows, particularly in time-constrained, high-acuity environments.

Our study extends these findings by demonstrating that even brief, opportunistic screening with a mobile single-lead ECG can identify a substantial proportion of previously undiagnosed AF cases in real-world clinical practice without requiring prolonged monitoring or structured protocols. However, exploratory comparison of automated classifications with chart-documented AF history showed only moderate concordance for identifying patients with known AF. These findings differ from diagnostic validation studies conducted under controlled conditions but should not be interpreted as formal estimates of sensitivity or specificity because a universal reference-standard ECG was not applied [10,17,18]. Moreover, no repeat measurements were mandated for unclassified tracing, and results were interpreted in real time by clinical staff without independent expert review. Consequently, our data reflects the raw utility of the tool in an elderly and acutely ill cohort highly prone to muscle tremors and baseline ECG anomalies. These figures serve as exploratory metrics rather than definitive diagnostic validation, as a universal 12-lead ECG gold-standard reference was not implemented for all participants.

Technical challenges were common; 16.0% (38/237) of recordings were inconclusive, with a higher rate in inpatients than outpatients. These findings reflect the practical realities of deploying digital tools in acutely ill or frail patients, who may have tremors, reduced mobility, or cognitive impairments. Patient acceptance of home-based screening was high (56.5%), but 42.6% declined, citing visual impairment, skepticism toward digital health, lack of interest, and concerns about technical difficulties. These barriers are not trivial. They reflect real-world disparities in digital literacy and access, particularly among older adults. Notably, only 42.5% of inpatients with AF were aware of their diagnosis. While most patients reported experiencing symptoms like palpitations or dizziness, only 1.5% could link them to AF. This profound gap between symptom awareness and disease recognition highlights a critical opportunity for patient education and digital health interventions.

Limitations

Several limitations must be considered. The study design may be susceptible to selection bias, as participants were recruited from high-risk inpatient and outpatient settings where AF prevalence is expected to be higher than in the general population. The authors were also not blinded to rhythm status. Real-time alerts from the KardiaMobile device were visible during screening, and confirmatory 12-lead ECGs were performed based on these results. This may introduce performance and detection bias. However, this study was designed as a pragmatic study, not a diagnostic validation study. The workflow reflects real-world practice, where clinicians are aware of patient context and use device alerts to guide care.

A further technical limitation was the inclusion of patients with permanent pacemakers or temporary pacing wires. In these patients, electronic pacing was frequently misinterpreted as signal artifacts, increasing the risk of false-positive AF detections or masking of underlying arrhythmias. Due to the absence of relevant data in the baseline dataset, these confounding factors could not be statistically adjusted for, potentially affecting the accuracy of the algorithm's performance assessment. Only inpatients from surgery, cardiology, internal medicine, and oncology departments were included, which may limit the generalizability of the results. The readings from the mobile device were not systematically cross-checked by an independent panel of physicians. This workflow was deliberately chosen to realistically reflect daily clinical routine, including its common time and staffing constraints. The absence of a universal 12-lead ECG reference standard and independent expert review means that diagnostic performance metrics should be interpreted as exploratory, not definitive. We also have to acknowledge that our 12-item questionnaire used in the study is not a validated scale and lacks the depth of formal qualitative interviews. However, we intentionally chose this approach as an exploratory questionnaire to provide a practical, descriptive assessment of what patients know and how ready they are for digital health. Instead of offering deep qualitative insights, this survey serves a hypothesis-generating purpose. By showing real-world knowledge gaps and physical barriers, these descriptive findings give future studies a clear starting point to use detailed interview methodologies.

Conclusions

This pragmatic feasibility study demonstrates that mobile single-lead ECG screening is feasible in routine inpatient and outpatient care and can be integrated into real-world clinical workflows. Technical challenges and barriers related to digital health literacy were common and should be considered in future implementation strategies. Because systematic confirmatory ECG testing was not performed in all participants, the study does not establish diagnostic accuracy, clinical benefit, stroke prevention, cost-effectiveness, or improved patient outcomes. Further prospective studies with rigorous validation and longitudinal follow-up are needed to evaluate the clinical impact and sustainability of mobile AF screening programs.

Acknowledgments

We would like to forward our deepest gratitude to Brandenburg Medical School and the outpatient and inpatient partners, supervisors, student data collectors, sponsoring organization, and study participants.

XSL•FO
RenderX

Funding

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The study was funded by the Brandenburg Medical School publication fund supported by the Ministry of Science, Research and Cultural Affairs of the State of Brandenburg.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Investigation and formal analysis: RF and DR
Oversight and leadership responsibility for the research activity: CB and AH-F
Conceptualization – ideas; formulation or evolution of overarching research goals and aims: AH-F
Writing original draft: AH-F, RF and DR
Creation and/or presentation of the published work, specifically visualization/data presentation: RF and DR
Data curation and validation: AH-F, JN, DM and FM
Resources – provision of study materials and patients: OR, SS, RM, MZ, ND and NL-K.
All authors reviewed and edited the manuscript.

Conflicts of Interest

SS is the editor-in-chief of *JMIR Cardio* at the time of this publication; however, SS was not in this role at the time of submission or review of this manuscript. SS had no involvement in the editorial review and processing of this manuscript. All other authors have no conflicts of interests to declare.

Multimedia Appendix 1
Questionnaire.
[DOCX File, 144 KB - cardio_v10i1e92799_app1.docx]

References

1. Elliott AD, Middeldorp ME, Van Gelder IC, Albert CM, Sanders P. Epidemiology and modifiable risk factors for atrial fibrillation. Nat Rev Cardiol 2023 Jun;20(6):404-417. [doi: 10.1038/s41569-022-00820-8] [Medline: 36600003]
2. Meyre P, Blum S, Berger S, et al. Risk of hospital admissions in patients with atrial fibrillation: a systematic review and meta-analysis. Can J Cardiol 2019 Oct;35(10):1332-1343. [doi: 10.1016/j.cjca.2019.05.024] [Medline: 31492491]
3. Krankenhäuser im Land Brandenburg 2023, Teil I Grunddaten. Statistik Berlin-Brandenburg. URL: https://www.statistik-berlin-brandenburg.de/a-iv-2-j [accessed 2026-08-28]
4. Krankenhausatlas. URL: https://krankenhausatlas.statistikportal.de/ [accessed 2024-11-17]
5. Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2024 Sep 29;45(36):3314-3414. [doi: 10.1093/eurheartj/ehae176] [Medline: 39210723]
6. National Provider Registry Statistics – regional distribution and density of specialized medical care in Germany. : Kassenärztliche Bundesvereinigung (KBV); 2023.
7. Gehrmann J, Hahn F, Stephan J, et al. Current use, challenges, barriers, and chances of telemedicine in the ambulatory sector in Germany-a survey study among practicing cardiologists, internists, and general practitioners. Telemed J E Health 2025 Jun;31(6):779-792. [doi: 10.1089/tmj.2024.0528] [Medline: 39909467]
8. Hall A, Mitchell ARJ, Wood L, Holland C. Effectiveness of a single lead AliveCor electrocardiogram application for the screening of atrial fibrillation: a systematic review. Medicine (Baltimore) 2020 Jul 24;99(30):e21388. [doi: 10.1097/MD.0000000000021388] [Medline: 32791751]
9. Andrade JG, Shah A, Godin R, et al. Cost-effectiveness of atrial fibrillation screening in Canadian community practice. Heart Rhythm O2 2023 Feb;4(2):103-110. [doi: 10.1016/j.hroo.2022.11.003] [Medline: 36873310]
10. Chan PH, Wong CK, Poh YC, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. J Am Heart Assoc 2016 Jul 21;5(7):e003428. [doi: 10.1161/JAHA.116.003428] [Medline: 27444506]
11. Hendrikx T, Sundqvist M, Sandström H, et al. Atrial fibrillation among patients under investigation for suspected obstructive sleep apnea. PLoS ONE 2017;12(2):e0171575. [doi: 10.1371/journal.pone.0171575] [Medline: 28178304]
12. Lowres N, Hillis GS, Gladman MA, et al. Self-monitoring for recurrence of secondary atrial fibrillation following non-cardiac surgery or acute illness: A pilot study. Int J Cardiol Heart Vasc 2020 Aug;29:100566. [doi: 10.1016/j.ijcha.2020.100566] [Medline: 32885031]

XSL•FO
RenderX

13. Smartphone-Nutzung in Deutschland – Daten & Fakten. Statista. URL: https://de.statista.com/themen/6137/smartphone-nutzung-in-deutschland [accessed 2025-07-21]
14. Freyer L, von Stülpnagel L, Spielbichler P, et al. Rationale and design of a digital trial using smartphones to detect subclinical atrial fibrillation in a population at risk: the eHealth-based bavarian alternative detection of Atrial Fibrillation (eBRAVE-AF) trial. Am Heart J 2021 Nov;241:26-34. [doi: 10.1016/j.ahj.2021.06.008] [Medline: 34252387]
15. Halcox JPJ, Wareham K, Cardew A, et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. Circulation 2017 Nov 7;136(19):1784-1794. [doi: 10.1161/CIRCULATIONAHA.117.030583] [Medline: 28851729]
16. Svennberg E, Friberg L, Frykman V, Al-Khalili F, Engdahl J, Rosenqvist M. Clinical outcomes in systematic screening for atrial fibrillation (STROKESTOP): a multicentre, parallel group, unmasked, randomised controlled trial. Lancet 2021 Oct 23;398(10310):1498-1506. [doi: 10.1016/S0140-6736(21)01637-8] [Medline: 34469764]
17. Koshy AN, Sajeev JK, Negishi K, et al. Accuracy of blinded clinician interpretation of single-lead smartphone electrocardiograms and a proposed clinical workflow. Am Heart J 2018 Nov;205:149-153. [doi: 10.1016/j.ahj.2018.08.001] [Medline: 30195576]
18. William AD, Kanbour M, Callahan T, et al. Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: the iREAD Study. Heart Rhythm 2018 Oct;15(10):1561-1565. [doi: 10.1016/j.hrthm.2018.06.037] [Medline: 30143448]

Abbreviations

AF: atrial fibrillation
CHA_2DS_2-VASc: congestive heart failure, hypertension, age, diabetes mellitus, prior stroke or TIA or thromboembolism, vascular disease, age, sex category
ECG: electrocardiography
FDA: US Food and Drug Administration

Edited by KC Wong; submitted 03.Feb.2026; peer-reviewed by J Claggett, S Turnbull; revised version received 24.Jun.2026; accepted 25.Jun.2026; published 10.Sep.2026.

Please cite as:
Fenske R, Rossa D, Butter C, Muehlensiepen F, Nübel J, Meretz D, Ritter O, Spethmann S, Mantke R, Zerbaum M, Draheim N, Lindenberg-Krause N, Haase-Fielitz A
Feasibility of Mobile Single-Lead Electrocardiogram Screening for Atrial Fibrillation: Multicenter Cross-Sectional Study
JMIR Cardio 2026;10:e92799
URL: https://cardio.jmir.org/2026/1/e92799
doi:10.2196/92799

© Robert Fenske, Denise Rossa, Christian Butter, Felix Muehlensiepen, Jonathan Nübel, Daniel Meretz, Oliver Ritter, Sebastian Spethmann, Rene Mantke, Mario Zerbaum, Nico Draheim, Nicole Lindenberg-Krause, Anja Haase-Fielitz. Originally published in JMIR Cardio (https://cardio.jmir.org), 10.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Involving Families in Cardiac Care Through Remote Patient and Family Management: Focus Group and Journey Mapping Study

Julian Houwen[1,2], MSc; Veronica R Janssen[1,3], PhD; Sara M Hondmann[3], MSc; Niels H Chavannes[4,5], MD, PhD; Maaike Kleinsmann[1,2,5], PhD; Douwe E Atsma[1,2,5], MD, PhD; Valeria Pannunzio[2], PhD

[1]Department of Cardiology, Leiden University Medical Center, Albinusdreef 2, Leiden, The Netherlands

[2]Department of Design, Organization and Strategy, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

[3]Unit of Health, Medical and Neuropsychology, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, The Netherlands

[4]Department of Public Health and Primary Care (PHEG), Leiden University Medical Center, Leiden, The Netherlands

[5]National eHealth Living Lab, Leiden University Medical Center, Leiden, The Netherlands

Corresponding Author:
Julian Houwen, MSc
Department of Cardiology, Leiden University Medical Center, Albinusdreef 2, Leiden, The Netherlands

Abstract

Background: In cardiovascular care, illness and recovery affect both patients and their families, particularly within home-based remote patient management (RPM). A recent scientific statement from the American Heart Association highlighted the importance of family involvement, identifying digital technologies as a key enabling opportunity. Despite this, research into the needs of families and the implications of RPM remains limited.

Objective: This study explored the lived experiences and unmet needs of patients with cardiovascular disease (CVD) and their relatives within RPM-supported cardiac care, using perioperative care and myocardial infarction as representative trajectories. Based on the identified gaps, we proposed a set of features for remote patient and family management (RPFM) interventions to address these needs.

Methods: This qualitative study was conducted at a Dutch university hospital with over a decade of experience in RPM across CVD pathways. A human-centered design approach was employed, including focus groups with 24 participants (13 patients with CVD and 11 relatives). Data analysis followed a framework analysis approach, combining inductive theme identification with deductive mapping onto existing frameworks. Care experiences and unmet needs were identified inductively and segmented along the Family Systems Illness Model's phases of illness. The needs were subsequently categorized deductively into the domains of the Supportive Care Framework. Based on these user-informed insights, RPFM features were generated through author ideation and internal team consensus. Finally, these experiences, needs, and features were synthesized into a visual journey map illustrating key care moments across 3 phases: preadmission, admission, and postadmission.

Results: We identified 47 unmet needs across 6 domains: informational (n=13), psychoemotional (n=13), social (n=7), physical (n=7), practical (n=6), and spiritual (n=1). The most significant gaps, described as "black holes" in care and support, emerged during the preoperative waiting period and early home recovery, which were characterized by a lack of information and psychoemotional support. To address the identified unmet needs, we generated 33 RPFM intervention feature ideas. These RPFM features were mapped across care phases and grouped into 7 categories: dynamic pathway navigation (n=8), on-demand support and information (n=5), family well-being modules (n=5), medical translation and consultation support (n=4), safety and assurance monitoring (n=4), collaborative lifestyle management (n=4), and peer support platform (n=3).

Conclusions: This study identified experience gaps in RPM-supported cardiac care. It showed that most unmet needs for care and support extend beyond hospital admission and discharge and that most health behaviors and recovery occur in the home context, within the patient's relational ecosystem with loved ones. The proposed RPFM features provide preliminary directions for exploration toward a transition from individually focused monitoring to inclusive, family-centered care and management.

(JMIR Cardio 2026;10:e83055) doi:10.2196/83055

KEYWORDS

family-centered care; cardiovascular disease; remote patient management; journey mapping; digital health; unmet needs

XSL•FO
RenderX

Introduction

Digital health technologies, including remote patient management (RPM), are increasingly embedded in cardiovascular care pathways with the goal of improving clinical outcomes, facilitating patient self-management, and decreasing pressure on the health care system [1-4]. RPM enables continuous guidance, real-time monitoring, and early clinical interventions while encouraging active patient engagement in self-care and recovery [5-8]. In turn, active patient engagement enables a range of opportunities for sustained lifestyle change, which significantly reduces cardiovascular risk [9].

Despite these developments and the growing emphasis on home-based self-care, most cardiovascular care models remain focused on in-hospital, staff-delivered care, underemphasizing the broader psychosocial and relational contexts in which cardiovascular illness is experienced [10,11]. As a result, the role of informal caregivers, such as family, is often overlooked [12]. However, these individuals play a crucial role in terms of cardiovascular risk management, home-based care, medication adherence, symptom monitoring, and emotional support [13]. Yet, they are often unrecognized in care delivery systems and therefore underprepared for their responsibilities, despite evidence of their impact on care quality and patient outcomes [10,12,13].

In a recent scientific statement, the American Heart Association has pointed out the importance of involving family systems in cardiovascular care [14]. Here the term “family” is defined in a broad sense, including not only biological relatives but also close friends, neighbors, or others whom the patient wishes to involve. The statement identifies digital technologies, including telehealth, as a key opportunity to support this integration [14]. This direction aligns with the goals of European health systems, where aging populations and health care staff shortages are increasing dependence on informal care [15,16].

Thus, there is now a clear need to advance RPM into remote patient and family management (RPFM), a model that actively includes and supports families throughout the care trajectory. As such, there is an urgent need to understand how RPM-supported models of cardiovascular care can effectively involve relatives throughout the care pathway, empower them in their role, and support them as individuals coping with the cardiac illness of a loved one.

Yet, empirical research on the needs of relatives, the implications of RPM for their caregiving experience, and what is required to design family-centered digital interventions that respond to their real needs remains limited. This study addresses this gap by exploring the lived experiences of patients and their relatives following RPM-supported cardiac care provided by a major university hospital in the Netherlands and by translating these experiences into RPFM intervention features.

In this study, we build on earlier research on family involvement in cardiac RPM [17] and employ a human-centered design approach that involves patients, family members, clinical experts, and design experts in a meaningful and iterative manner throughout the research process [18]. Using focus groups and journey mapping [19-24], we identify user-informed key care moments, unmet needs, and opportunities for family-centered RPM. Our analytical framework draws on the Family Systems Illness Model [25] and the Supportive Care Framework [26] to situate informal caregiving within a broader relational and psychosocial context.

By studying these often-overlooked family experiences and designing meaningful solutions, we aim to contribute to the development of more inclusive, “real-life” RPM-enabled cardiovascular care pathways that support both patients and their relatives in effectively coping with the disease trajectory and lifestyle modification. In doing so, we aim to contribute to improving both experiential and health outcomes for patients and their loved ones.

Methods

Setting and Design

This qualitative study was conducted at a university hospital in the Netherlands, with over a decade of experience in RPM [8,27-31]. It explored the experiences of patients and their relatives in 2 distinct patient groups, selected as representative examples of cardiovascular disease (CVD) care pathways: acute myocardial infarction (MI) and elective cardiac surgery within a perioperative (peri-OP) care pathway. These care pathways, despite differences in clinical presentation, were selected to provide broad, varied insight into the experiences of the largest cardiology patient populations and their relatives. In addition, both are supported by the same RPM program, *The Box* [8,27-31], which enables hospital service-level comparison. *The Box* offers hybrid care using eHealth devices (eg, smartwatch, blood pressure monitor, digital scale, and thermometer) and a connected patient-facing mobile app. The service is currently used for remote home monitoring of patients’ vital parameters, enabling clinical oversight and timely intervention while reducing the need for in-person hospital visits. Now that the “monitoring” foundation is in place in hospital practice, a shift in the intervention toward “self-management” has been initiated. This study contributes to this iteration by offering recommendations that broaden the service functionalities through involving and integrating patients’ family environments.

To develop recommendations for RPFM interventions, a human-centered design approach was employed. The process involved identifying patient and family experiences through focus groups, from which user-informed unmet needs were extracted. These needs then informed the generation of RPFM intervention features through author ideation and internal team consensus. Finally, the results were synthesized into a visual journey map. As a last step, the journey map and features were shared with participants via email for validation; however, no feedback was received. The study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines for qualitative research reporting [32].

Participants and Recruitment

Patients in the MI and peri-OP pathways who were ≥2 months after hospital admission and still in hospital-based follow-up through the hybrid care program were purposively sampled.

XSL•FO
RenderX

We aimed for a sample size sufficient to conduct at least 6 focus groups because this amount was found to uncover 90% of relevant themes in previous research [33]. Participants were invited via email, followed by telephone confirmation. Participants were encouraged to bring a family member, caregiver, neighbor or friend closely involved in their care. Recruitment aimed for variation in age, gender, and caregiving relationships. There were no preexisting relationships between the researchers collecting the data and the participants.

Focus Groups

Seven focus group sessions were conducted at the hospital in June 2024, divided by care pathway. In 6 sessions, patients and relatives participated in separate groups to enable open discussion. One session combined both groups for practical reasons, allowing complementary patient-relative perspectives. Group sizes ranged from 3 to 4 participants per session (see Multimedia Appendix 1 for an overview). No significant differences in participants' candor or the depth of shared experiences were observed between the combined and separate formats.

Each focus group session lasted approximately 2 hours. Every session ended with an interview wrap-up to determine whether, according to the participants, the most important topics and points were covered. Moderation was carried out by 2 researchers (JH, a male PhD researcher and strategic designer, and SMH, a female PhD candidate and medical psychologist), supported by a female medical psychology intern.

Semistructured discussions were held around the "Care and Well-Being Journey" template (available in Multimedia Appendix 2). This is a participatory mapping tool designed for this study based on existing journey mapping approaches [19]. Discussions covered 3 phases: preadmission, admission, and postadmission. Participants discussed key moments, challenges, and support experiences across the need domains of the Supportive Care Framework [26]. Post-it notes were used to annotate experiences on the template.

Data Analysis

The discussions were recorded, and the audio was transcribed. The transcripts were analyzed using a framework analysis approach [34], combining inductive theme identification with deductive mapping onto existing frameworks and adapted for the purpose of journey mapping. The analysis aimed to generate themes related to participants' care experiences and unmet needs, which were later mapped onto a patient journey format [19,20]. To structure the analysis, we used 2 theoretical models: the Family Systems Illness Model's phases of illness [25], which informed segmentation of family experiences across the care pathway, and the Supportive Care Framework [26], which provided domains for categorizing unmet needs.

The analysis was performed in 3 steps. First, after familiarization with the transcripts, coding was conducted inductively by 2 researchers (JH and VP, a female assistant professor). They assigned meaningful labels to transcript text segments related to care experiences and unmet needs while simultaneously tagging each code with its phase along the cardiac care pathway. This step was conducted in ATLAS.ti 2025 (ATLAS.ti GmbH) software. Second, the codes were clustered and refined across multiple rounds of discussion, moving from initial codes to broader themes. These themes were visually organized into a journey map structure using Miro (RealtimeBoard, Inc) software. Third, within each theme, specific unmet needs for patients, relatives, and families were deductively categorized into the domains of the Supportive Care Framework [26]. To better reflect the data, we made a minor adaptation to Fitch [26] domains: psychological and emotional needs were combined into a single psychoemotional category, as participants did not seem to distinguish between these aspects during discussion.

No formal data saturation assessment was performed. Instead, the process continued through multiple rounds of consensus building until the research team agreed that the themes accurately reflected the participants' experiences. Any discrepancies were resolved through discussion. Throughout the analysis, the researchers engaged in ongoing reflexivity, using transdisciplinary team discussions to examine how their backgrounds in medicine and design influenced theme development and to mitigate individual professional biases.

RPFM Features Ideation

The themes and corresponding unmet needs were used by the first author (JH) to ideate RPFM intervention features. These were iteratively refined through internal multidisciplinary team consensus discussions involving clinical experts (VRJ, a female associate professor and medical psychologist, and DEA, a male professor of cardiology) and design experts (VP and MK, a female professor of design for digital transformation). This design process ensured that the features were grounded in the lived experiences of families and aligned with clinical and psychosocial realities.

Journey Mapping

Finally, the collected needs and corresponding RPFM feature ideas were mapped visually, with the map iterated through multiple rounds to improve consistency and readability. This journey map offers both an overview of unmet needs across patients' and families' current journeys and a summary of additional RPFM features that could improve specific touchpoints.

Ethical Considerations

Ethical approval was granted by the institutional nWMO-div1 review board at the Leiden University Medical Center (DAT/tak/1002024). Written informed consent was obtained from all participants prior to participation. All collected data were deidentified. As a gesture of appreciation for their participation, each participant received a €15 (approximately US $16) gift card.

Results

Participant Characteristics

A total of 42 individuals were invited to participate in the study, comprising 22 patients with CVD and 20 relatives. Of these, 9 pairs of patients and relatives did not participate, citing reasons such as illness, forgetting the scheduled session, or providing no explanation. This resulted in the inclusion of 24 participants

XSL•FO
RenderX

(13 patients with CVD and 11 relatives). All participants received care and support primarily from the same university medical center and affiliated rehabilitation center. An overview of participants' characteristics is presented in Table 1.

Table . Demographics and clinical characteristics of patients with CVD[a] and relatives (N=24) in the qualitative focus groups.

Characteristic	Patients with CVD (n=13)	Relatives (n=11)	Total (N=24)
Sex, n (%)			
Female	8 (62)	5 (45)	13 (54)
Male	5 (38)	6 (55)	11 (46)
CVD care path, n (%)			
Peri-OP[b]	6 (46)	5 (45)	11 (46)
MI[c]	7 (54)	6 (55)	13 (54)
Relation to patient, n			
Partner	—[d]	7	7
Child	—	2	2
Parent	—	2	2
Age group (years), mean (SD; range)[e]	60.8 (12.6; 30 - 80)	57.3 (17.7; 20 - 85)	59.2 (14.9; 20 - 85)

[a]CVD: cardiovascular disease.

[b]Peri-OP: perioperative.

[c]MI: myocardial infarction.

[d]Not applicable.

[e]Mean and SD were estimated using the midpoints of the reported age ranges.

Journey Map

Through our human-centered design approach, we mapped the real-life experiences of individuals and their families dealing with cardiac events and symptoms, cardiovascular procedures, and their aftermath. Based on the data from the focus group sessions, we distilled a set of unmet needs across these phases, which we then used to generate RPFM interventions through author ideation and internal team consensus. The results of both steps are presented in a comprehensive journey map that displays categorized needs and related RPFM ideas.

The full journey map is shown separately in Figures 1-3, visualizing a unified care pathway for MI and peri-OP care. Figure 1 illustrates the preadmission phase, including symptoms, diagnosis, waiting time, and the admission phase, spanning preprocedure to postprocedure treatment and discharge. Figure 2 presents postadmission with phase-specifics covering recovery at home, rehabilitation, and long-term management. Figure 3 also addresses the postadmission period, focusing on longitudinal experiences that could not be situated within a defined timeframe. In addition, within and across these phases, we identified unmet needs and presented them by experience themes, specifying whether they apply to patients, relatives, or families as a whole. These are color-coded by need domain (adapted from Fitch [26] Supportive Care Framework; see *Methods*). Additionally, RPFM intervention features to address these unmet needs are mapped in the journey (indicated by icons) under the related unmet needs.

XSL•FO
RenderX

Figure 1. Journey map of the remote patient management–supported cardiovascular disease care pathway part 1: key phases, patient and relative experiences, unmet needs, and proposed RPFM intervention features. MI: myocardial infarction; peri-OP: perioperative; RPFM: remote patient and family management.

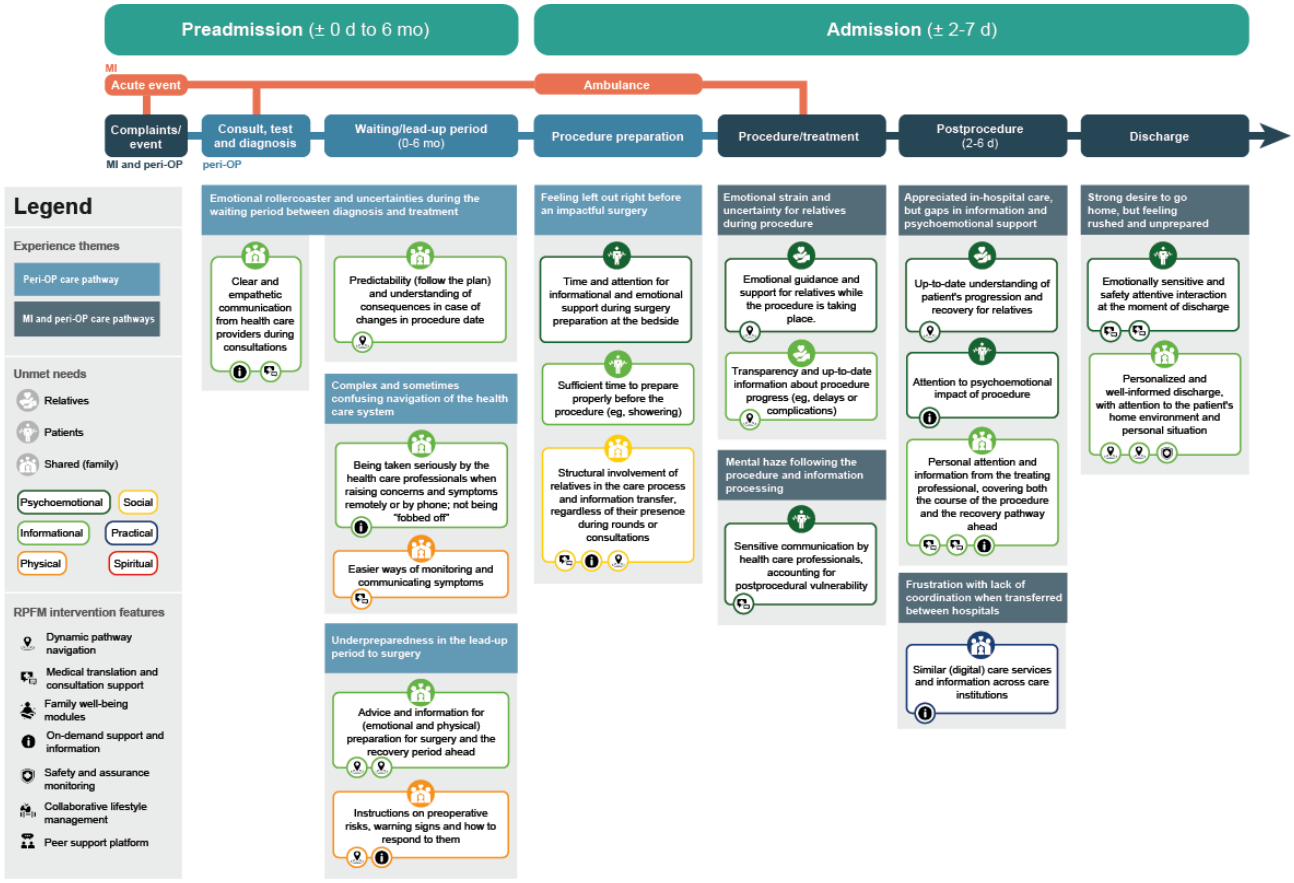
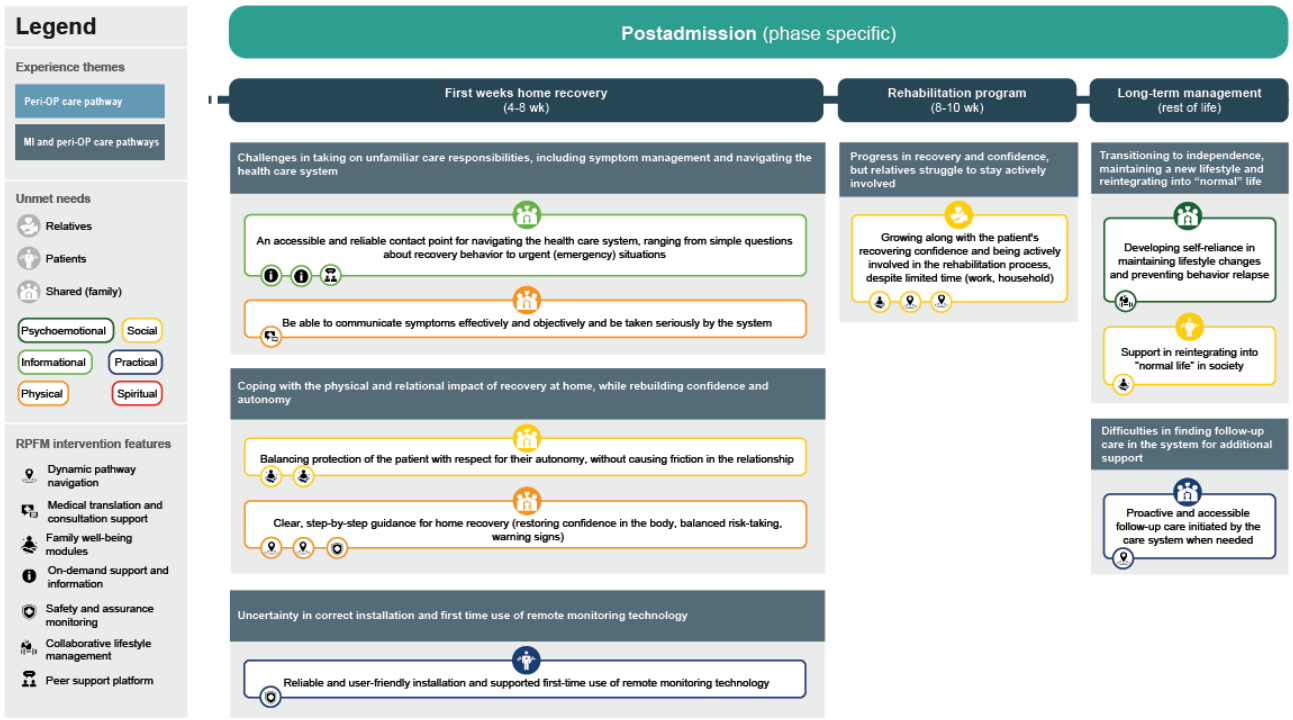
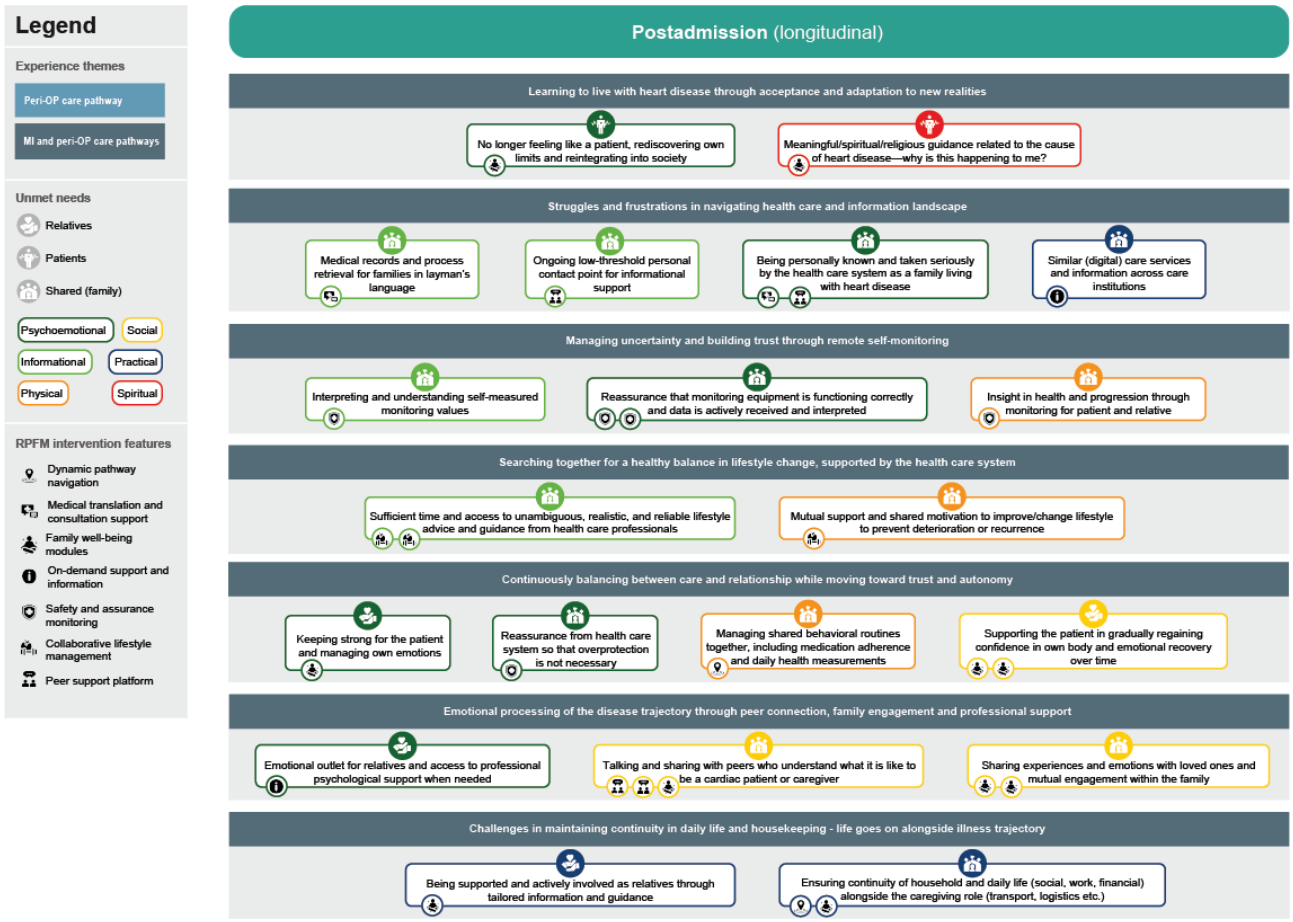


Figure 2. Journey map of the remote patient management–supported cardiovascular disease care pathway part 2: key phases, patient and relative experiences, unmet needs, and proposed RPFM intervention features. MI: myocardial infarction; peri-OP: perioperative; RPFM: remote patient and family management.



XSL•FO
RenderX

Figure 3. Journey map of the remote patient management–supported cardiovascular disease care pathway part 3: key phases, patient and relative experiences, unmet needs, and proposed RPFM intervention features. MI: myocardial infarction; peri-OP: perioperative; RPFM: remote patient and family management.



Unmet Needs and RPFM Feature Ideas

Overview

Our analysis revealed a broad range of unmet needs among patients, relatives, and families across different phases of the care pathways. A total of 47 unmet needs were identified and displayed in the journey map. These reflect significant gaps extending well beyond clinical management to the informational (n=13), psychoemotional (n=13), social (n=7), physical (n=7), practical (n=6), and spiritual (n=1) domains. Regarding the experienced impact of unmet needs, the most important phases were the "waiting/lead-up period" in preadmission (specifically peri-OP care) and the "first weeks of home rehabilitation" postadmission. These periods were considered by several participants to be "black holes" in care and support. During the admission, the unmet needs were experienced as less urgent, and clinical care was particularly praised.

Furthermore, the generation of human-centered ideas for RPFM features yielded 33 features applicable across the journey. These features were grouped into 7 main categories (see Table 2). A more detailed overview is available in Multimedia Appendix 3.

Some interventions are repeated throughout the journey as they respond to needs across multiple phases. These features are intended to enhance family-centered care by actively empowering family participation, improving information transmission, assisting navigation within health systems, aligning expectations, and providing holistic support throughout the cardiac care journey for patients and relatives.

Below is a detailed presentation of the key unmet needs, categorized by domain, along with examples of corresponding ideated RPFM features that may offer solutions.

XSL•FO
RenderX

Table . Ideated remote patient and family management (RPFM) intervention categories for patients with cardiovascular disease and relatives: descriptions, core focus, and RPFM intervention features.

Number	RPFM intervention categories	Core focus	Ideated RPFM intervention features
1	Dynamic pathway navigation (n=8)	What to do? What can we expect? How to do it?	• Personalized recovery trajectory and goal planner • Real-time guided expectation management • Shared medical routine and event planner • Joint health overview and progression • Relatives' rehabilitation insight feed • Hospitalization support guidance • Risk and symptom management instructions • Follow-up care navigator
2	On-demand support and information (n=5)	Help us (now), please?	• Emotional support and crisis access • Digital emergency assistant • Integrated family care-portal • Intelligent information-assistance • Knowledge navigator
3	Family well-being modules (n=5)	How do we cope and deal?	• Social and practical support guide • Caregiving guidebook • Self-assessment check-ins • Family dialogue toolkit • Psychosocial resilience modules
4	Medical translation and consultation support (n=4)	Do we (families and the care system) understand each other?	• Symptom and feeling translator • Question logbook • Family attendance scheduler • Plain-language consultation archive
5	Safety and assurance monitoring (n=4)	Is it going well? Is it safe?	• Discharge checklist • Installation and connection validation • Data-driven personalized feedback • Automated monitoring feedback loop
6	Collaborative lifestyle management (n=4)	What can we do to prevent problems, and how do we get better?	• Lifestyle information database • Balanced and personalized lifestyle coaching • Interpersonal motivational messaging • Lifestyle assessment and feedback
7	Peer support platform (n=3)	Who do really understand us?	• Buddy matching service • Online peer support forum • Virtual group sessions

Informational

Perioperatively, families reported inadequate communication from health care providers about the care process. They experienced uncertainty due to insufficient information (eg, about changes in procedure schedules and their impact) and felt that they were not taken seriously and were dismissed without clear explanations when expressing their concerns and questions. Guidance on behavioral, emotional, and physical preparation

XSL•FO
RenderX

for the medical procedure and recovery was perceived as inadequate, with some families feeling rushed and unprepared due to limited consultation time and unclear instructions on expected behaviors and protocols. Family members also missed timely, transparent updates on the procedure's progress, especially in cases of delays or complications.

In the preoperative phase, RPFM could support families in preparing for the unknown and reducing anxiety by providing clear information about the diagnosis and the entire care pathway. This could include summarized reviews of consultations and guided navigation to trusted sources. It may also offer preparation guidance and expectation management materials, such as videos explaining what to expect and advice on mutual physical and psychological readiness for the procedure, including lifestyle considerations. A personal care pathway tracker could offer families an overview of the care process. To further empower families to navigate the health care system, care navigation support could be included, for example, in the form of a decision tree indicating where and when to ask specific questions. On the day of the procedure, RPFM might deliver real-time digital updates in the event of delays or complications; for instance, if the procedure takes more time, a digital notification could be sent to the waiting family.

Postoperatively, patients missed direct and timely contact with their surgeon to discuss the process and outcome of the procedure. Families required more personalized information about the expected impact of heart disease on their care pathway and lives. Discharge from the hospital to home was often perceived as rushed, and the patient's environment and circumstances were not sufficiently considered. At home, there was no easily accessible contact person for questions and support, nor was there clear guidance on navigating the health care system, especially in the case of alarming symptoms. Families also experienced problems accessing medical records from different organizations and needed explanations to understand them. Time to discuss lifestyle changes for recovery was often perceived as limited, with insufficient personalized advice and guidance.

In this postoperative phase, RPFM could support families with expectation management on both physical and mental recovery, alongside a discharge checklist and advice tailored to the home situation. Instead of receiving a confusing stack of papers, families could be provided with a personalized digital recovery plan outlining clear next steps and offering access to a digital Q&A system for immediate concerns. The platform could also offer a summary of the procedure's course to clarify the patient's "lost hours," which is the period during which the patient was under anesthesia, and provide a step-by-step guide to the recovery pathway. For ongoing support, RPFM might feature a digital emergency assistant and facilitate access to low-threshold advice for informal guidance through a "buddy" system. It could also support personalized and balanced lifestyle coaching, facilitated by assessment tools and a trustworthy information database. Finally, RPFM could provide explanations of the collected data and offer a retrievable medical dossier summary in accessible language, empowering families to better understand the patient's health status.

Psychoemotional

Families felt inadequately prepared for the care process due to limited consultation time and a lack of guidance regarding the emotional impact of the procedure. Both patients and their relatives experienced increased anxiety, stress, and frustration due to unclear or insufficiently empathetic communication by health care providers, especially during key moments such as consultation, hospitalization, discharge, and early recovery. There was also a sense of uncertainty in the reliability of the home monitoring systems due to limited feedback and doubt about whether data were being actively monitored by professionals.

Family members indicated that involvement in updates on the patient's condition was limited, indirect, and not tailored to their emotional needs. During the recovery phase, both patients and family members struggled with the emotional impact of the illness and the effort to regain a sense of normality. Patients struggled to cope with "being a patient," and with regaining emotional stability and autonomy. Family members felt inadequately prepared to support the coping process and reported a lack of accessible emotional support or structured channels to express their own fears, doubts, and emotional burden.

Patients and family members did not distinguish between psychological and emotional care. However, the analysis revealed a spectrum of illness-related distress. This ranged from general situational worries to potential clinical "red flags," such as prolonged depressive thoughts and structural fears of recurrence when being alone. Discussing these needs even motivated a participant to express the need for additional mental health care from the hospital following the focus group session.

We propose using an RPFM approach to address these unmet psychoemotional needs, depending on severity. An example of more general support is providing digital updates to family members in an empathetic manner (content, tone of voice, and timing) that considers the stressful situation and emotional state, as well as providing guidance during impactful moments such as surgery. To build trust and reduce distress related to monitoring, RPFM could offer automatic reassurance notifications and proactive feedback, supported by personal data analysis. For more intensive support needs, it could provide direct emotional support avenues (eg, a crisis helpline), integrate individual emotional processing tools for relatives, and feature a mental health check. For broader coping and support, RPFM could facilitate peer connection through modules on living with chronic conditions, buddy systems, and shared experiences.

Social

Families reported insufficient opportunities for active involvement of relatives throughout the care journey and during information transfer. Relatives lacked support in balancing psychosocial care with patient autonomy and struggled to stay engaged in the care process due to their own responsibilities and other constraints. Relatives felt unequipped to assist patients and themselves in regaining trust in the body and in physical and emotional recovery and wished to be more involved in care. Support to adopt a healthier lifestyle together was often lacking, as was guidance on how to help the patient without causing

XSL•FO
RenderX

friction. Also, the opportunity for family members to connect with peers for emotional or experience exchange was not accessible or structurally arranged.

To support greater social involvement and role clarity, RPFM could provide realistic expectation management for patients and relatives regarding active participation during admission. To navigate the delicate balance between support and patient autonomy, it could offer communication advice (eg, "How do I talk about this?") and modules on discovering new boundaries for both patients and relatives (eg, a "Care & Release" coaching module that could support caregivers in transitioning from a period of intensive caregiving back to a more supportive role, allowing the patient to regain autonomy). RPFM could also include patient support training and parallel rehabilitation modules specifically for relatives, equipping them to provide effective, shared support. For sustained engagement and community building, RPFM could integrate an online peer support or matching platform for group or individual discussions for both patients and relatives, along with guidance on how to foster mutual support within the family unit.

Physical

Patients and families lacked concrete, step-by-step guidance for preparing for the medical procedure and home recovery, including support in restoring confidence in the body, safely weighing risk-taking, and recognizing and responding to warning signs. There was a need for simple, user-friendly tools to enable understanding, tracking, and communication of symptoms to health care providers. Furthermore, families required a shared understanding of recovery progress, medication adherence, and the patient's emotional state to support understanding and confidence in both the body and the care process.

To meet these physical needs, RPFM could provide concrete guidelines for families to monitor and manage peri-OP and postoperative risks and symptoms, coupled with easy access to reliable information and a symptom "explainer" (eg, a pain/emotion score) for clear communication with the health care system. The platform could integrate rehabilitation modules and a personalized recovery planner, leveraging data to ensure that patients and families have a clear, shared understanding of the recovery journey. Furthermore, it could increase engagement through motivation modules and support for setting realistic, joint lifestyle change goals while also offering a shared health overview providing insight into progress and routines on medication or behavior change adherence.

Practical

Patients and relatives experienced practical issues due to insufficient data interoperability between institutions, which led to duplication and a lack of uniformity in care processes. Digital care services, such as the patient portal and installation, were frequently experienced as complicated or inconsistent, obstructing effective use. Relatives lacked remote access to touchpoints and information when they could not attend consultations due to work or household commitments. The provision of accessible, proactive follow-up care was perceived as inadequate. Additionally, there was limited practical support for accessing social services, for example, for maintaining daily household routines and managing finances alongside the ongoing care.

To address these practical challenges, RPFM could streamline care by providing a unified, user-friendly digital platform. It could offer a family-centered hub that ensures consistent information and centralized data storage and exchange. For instance, if a relative misses a consultation due to work, RPFM could offer remote access to summaries and a logbook for questions, empowering them to stay involved. It could also enable proactive care planning, with features such as a shared routine planner. Additionally, it could integrate digital guidance for social services and navigation of follow-up care in the region, easing the burden of managing daily life alongside care.

Spiritual

Finally, patients mentioned the lack of existential, religious, or spiritual guidance in processing and coping with the impact of the condition, a dimension that received little attention in the care pathway.

RPFM could support this through dedicated modules designed to offer guidance on dealing with existential questions and finding meaning.

Discussion

Principal Findings

Our study provides insights into the lived experiences of patients and their relatives following MI or peri-OP RPM-supported care pathways in a university hospital in the Netherlands. Through our human-centered design approach, we identified unmet psychoemotional, informational, social, practical, physical, and spiritual needs for both patients and their relatives. These results align with previous studies on family involvement in cardiovascular care [35-39] and deepen previous knowledge by detailing lived experiences across a broad spectrum of needs. Particularly, our results reinforce the importance of both improved information provision and improved psychological and emotional support to patients and family members. To achieve both goals, we recommend several RPFM intervention feature ideas for the current RPM-supported care pathways. In doing so, we directly address a research gap identified by Goldfarb et al [14], who highlight the potential of health technology to advance family engagement in cardiovascular care. It is important to note that the generated feature ideas, in their current state, are hypothetical, unvalidated prototypes rather than a finalized set of solutions ready for implementation.

Our results indicate that 2 stages in the pathway were particularly undersupported: the period prior to admission (for the peri-OP pathway) and the early recovery phase at home (for both pathways). These phases were described by patients and family members as "black holes" in care and were specifically characterized by a lack of information and psychoemotional support. These findings are consistent with previous research reporting similar challenges during the peri-OP and early recovery phases [40-43]. Medical professionals often focus on optimizing in-person encounters, such as consultations and admissions. However, we identified that, from a family

XSL•FO
RenderX

perspective, the largest gaps in support and care occur during the extended periods spent at home. The missed support during these phases imposes a considerable burden on both patients and their relatives. Families sometimes feel abandoned and alone when it comes to managing fluctuating symptoms, taking over care responsibilities, and navigating the health care system. These challenges are intensified by the emotional impact of coping with heart disease while attempting to reestablish normalcy in daily life.

Comparison to Prior Work and Implications

Illness and care cannot be separated from everyday life, in which patients live together with their loved ones in a caring ecosystem [25]. Most care, health behavior, and recovery take place outside professional settings [10,44]. It is, therefore, essential that the health care system extends its reach and actively supports these everyday contexts. This is especially important as hybrid home-based care models shift more care responsibility to patients and their family members. In that light, our results align with and expand upon the findings of Schmid et al [45]. They observed that peri-OP education is foundational in shaping cardiac patients' subsequent behaviors and expectations regarding their treatment, particularly with respect to physical activity, nutrition, and mental health. Furthermore, they found that active patient and family engagement in the treatment process, including physiotherapy and mental support, significantly improves postoperative care and leads to faster recovery.

The proposed RPFM approach may represent a promising direction for addressing the identified gaps in care experiences, though it requires empirical validation. It offers hypothetical opportunities for continuous support and tailored interventions to be delivered directly into living environments. In this way, RPFM can reach patients and relatives at the moments when the "black holes" in information and support currently occur. Notably, these "black hole" phases overlap with the evidence-based window of opportunity where patients are most receptive to behavior change, as described in the framework of teachable moments [46]. Strategically supporting and educating families with RPFM interventions during these highly receptive periods could serve as a dual prevention mechanism, providing secondary prevention for patients and primary prevention for family members. This is particularly relevant as relatives may have similar risk profiles due to a common lifestyle, genetic predisposition, or shared environmental factors.

It is important to note that the identified unmet needs are not just "soft," medically irrelevant experiential issues but have direct clinical and health system implications [47,48]. In this regard, we build on previous RPM research that describes the interconnectedness of patients' experiences with other outcome measures [49]. By addressing the unmet needs in care and support reported by our participants, we hypothesize that the proposed RPFM approach could lead to outcome improvements across the components of the Quadruple Aim: health outcomes, patient experience, staff experience, and cost-effectiveness [50].

For instance, regarding health outcomes and cost-effectiveness, empowering families to recognize symptoms could improve clinical safety. For example, better-informed families might help identify early warning signs, which could theoretically reduce postoperative adverse events or prevent unplanned procedural postponements. Furthermore, RPFM's ability to support joint remote management and lifestyle changes may also extend its benefits beyond the patient, promoting primary prevention for family members who may have similar risk profiles as the patients.

As for patient and family experience, RPFM could possibly directly address identified informational and psychological gaps. Tailored digital information on preoperative, postdischarge care, and recovery trajectories may enhance knowledge and navigation of care. To meet psychological needs, integrating therapeutic modules could address and support mental health. Through such features, RPFM could hypothetically help to mitigate anxiety, posttraumatic stress disorder, and depression, issues that, at times, affect both cardiac patients and their relatives [51,52]. These interventions may also reduce anxiety-driven overprotectiveness, which can hinder rehabilitation [53]. However, empirical testing is needed to establish whether such features would achieve these outcomes.

Finally, improving patient and family experience could directly impact staff experience. By reducing miscommunication and streamlining inquiries and support processes, RPFM could decrease staff workload, limiting repetitive inquiries and administrative friction.

Future Directions

Overall, we propose shifting from home-based cardiac RPM to an RPFM approach that builds on the existing RPM infrastructure while also expanding it to explicitly integrate the family. This may include educational content, emotional support tools, and role-specific guidance tailored to the experiences of family members. This new paradigm acknowledges the relational dynamics of illness and recovery in home-based care models. Effectively involving families and meeting their needs can improve psychological health, self-efficacy, caregiving, quality of life, social support, and problem-coping skills. This, in turn, may lead to improved self-care and reduced hospital readmissions and unnecessary clinical visits [54]. However, empirical research is required to establish the impact and cost-effectiveness of the proposed RPFM interventions.

From a systems perspective, it is noteworthy that the 2 identified "black holes" in care and support needs (the preoperative waiting period and early home recovery) appear to align with the existing divisions between specialized hospital care and general practitioner care in the Dutch care system. At these transition points, families must navigate between different care environments while the breadth of support needs remains (eg, managing symptoms, seeking guidance, and balancing emotional and practical concerns). This occurs across care systems with different structures, protocols, and points of contact, contributing to the uncertainty already evident in these phases.

Here, an opportunity exists for RPFM to act as a connecting layer between the hospital and primary care by increasing the mobilization of informal family systems, whose involvement is constant throughout the different phases of the care journey. However, creating a coherent RPFM care pathway that is clear

XSL•FO
RenderX

to patients and families, with clear guidance on where to seek care and support at each step, will require cross-boundary collaboration and integration between hospital and primary care providers. Future research should therefore actively develop and test RPFM-supported care pathways that operate seamlessly across care lines and engage multiple types of health care professionals. Furthermore, it is important to investigate the enabling conditions for this, for example, insurer reimbursement and data interoperability.

Finally, reflecting on our human-centered design approach and identified opportunities, we propose an overall strategy for health care innovators wanting to design RPFM interventions. The strategy is presented in Figure 4. It offers a set of actionable steps, starting with an in-depth, empathetic understanding of patients and families through focus group discussions and journey mapping and concluding with the scaling of effective interventions across other care pathways. This overview also illustrates our plans for future RPFM research in relation to The Box [8,27-31], including the steps from 5 to 9.

Figure 4. A proposed 9-step future innovation strategy for designing remote patient and family management interventions. RPFM: remote patient and family management.



To operationalize this vision in practice, we aim to supplement and expand the current "The Box" service into a "Family Box," provided that the proposed prototype features prove to be feasible and desirable after thorough testing with users. Specifically, this would involve developing a family member-facing version of the existing smartphone app (with its own login and content), in which to implement validated, personalized RPFM features. In doing so, it will be essential to prevent digital overwhelm. Patients and their loved ones recovering from and coping with cardiac events already experience cognitive fatigue and psychoemotional distress, reducing their capacity for information processing. Introducing an extensive package of digital features risks increasing this burden rather than alleviating it. Therefore, the 33 proposed features should not be implemented simultaneously but rather prioritized and bundled toward a minimal viable product-service-system. This should be developed iteratively through active co-design with patients, families, and domain experts. The features could also be delivered in a modular fashion, aligned with the patient's and family's stage in the care journey, ensuring that only contextually relevant support is offered at each phase. This approach to modular, journey-aligned service delivery has been explored in our previous work on family involvement in RPM-supported cardiovascular care [17],

XSL•FO
RenderX

which identified heterogeneous family contexts and the need for personalized, adaptive interventions over time. Ideally, these modules would be integrated into a single digital interface to provide a seamless digital experience for both patients and their relatives, ensuring that the RPFM intervention is technically feasible, accessible, and genuinely supportive, rather than adding cognitive burden on recovering patients and their families. Beyond updates to the app, our recommendations include hospital service-level shifts, such as integrating relatives into clinical (tele-)consultation workflows and ensuring that families are included in active information provision, education, and psychosocial care throughout the entire care pathway.

While the proposed strategy originated in cardiovascular care, we find that it could be transferable to other clinical trajectories and disease areas. By following this strategy, innovators may find new opportunities for RPFM solutions that address real-life needs and achieve positive societal impact for families and health care systems.

Strengths and Limitations

A strength of our approach lies in the integration of families' lived experiences into a structured journey map, in alignment with Zøylner et al [55] methodology for mapping patient and relative experiences. In addition, our study expands on existing family journey mapping methodology by integrating theoretical contributions from the Family Systems Illness Model [25] and the Supportive Care Framework [26], embedded in the focus group sessions through the *Care and Well-Being Journey* template (Multimedia Appendix 2). In this way, we could distinguish between domains of family needs in a structured, theory-informed fashion. The proposed features ideated in this study offer starting points for clinicians and innovators seeking to embed family-centered RPM-supported care within cardiac care pathways. However, these are preliminary ideas, based on assumptions that require structured testing. Future research should include concept testing, pilot implementation, health economic evaluations, and studies to measure the impact on both patient-reported and family-reported outcomes and experiences. Another strength of our approach lies in its applicability to care domains other than cardiology, especially to other acute and chronic care pathways where long-term management and self-care take place at home, such as in the fields of oncology, nephrology, and obstetrics.

This study has several limitations. First, a "well-patient" and "super-user" selection bias may have affected the findings, as the sample consisted of native Dutch participants recruited from an existing RPM program (The Box), who had been enrolled for at least 2 months. Despite active and valued participation in the focus groups, participants did not react to a later chance to validate the results, which may be due to the nonbinding and possibly high-threshold nature of requesting feedback via email. The nonresponse constrains the reliability of the consensus on the ideated RPFM intervention features.

Additionally, there was a high nonparticipation rate (roughly 43%) due to reasons including illness. Consequently, the sample likely represented healthier, more engaged families, potentially introducing a risk of selection bias toward digitally literate families. This may have obscured the unmet needs of more vulnerable families who were too ill to participate or lacked the digital skills to register in the first place. Moreover, we did not collect data on socioeconomic status, educational attainment, or baseline digital literacy. Given that the proposed feature ideas are predominantly digital, this represents a significant limitation, as the proposed solution directions may prove inappropriate or inaccessible for lower-income or less digitally literate populations.

Furthermore, while this study primarily focuses on digital opportunities and on contributing to the iteration of the current version of an RPM service, the identified unmet needs might also be addressed through nondigital solutions. Examples are simplified information resources or personal, specialized extra care case managers, which may be more appropriate for less digitally literate families. Future research should intentionally include these vulnerable groups to explore alternative solution directions and ensure that RPFM interventions do not increase health disparities.

Second, there may be a recall bias that influenced the data collection, given the retrospective nature of the journey mapping. As stated before, patients were recruited at least 2 months after admission, yet they were asked to map the "Acute event" and "Admission" phases. As memory of the acute phase is often fragmented in trauma and cardiac care, we should acknowledge that reported "informational needs" during the acute phase might reflect a failure of retention due to stress rather than a failure of provision.

Third, this study merged the informational and psychological experience trajectories of the MI and peri-OP care pathways onto a single journey map. While both pathways are supported by the same RPM program, their clinical presentations and trajectories differ throughout the journey. Most notably, the peri-OP pathway involves an extended preoperative waiting period that is absent in the acute MI trajectory. Additionally, within the peri-OP pathway itself, postadmission recovery trajectories may differ considerably depending on the type of procedure, as wound healing and physical recovery following open-heart surgery differ from those following minimally invasive interventions. Merging these pathways and types of procedures onto a unified journey may have obscured pathway- and procedure-specific nuances, potentially limiting the clinical specificity and applicability of the recommendations. Future research should therefore consider testing and validating the ideated RPFM features in a pathway- and procedure-specific way, as the relevance, design, and implementation of features may differ case by case.

Fourth, the focus group sessions were small, ranging from 3 to 4 participants per session. Focus group sizes in the literature often range from 6 to 12 participants [33], and our sessions therefore fell below this recommended range. These small group sizes may have limited the interactive and synergistic dynamics characteristic of focus group methodology, potentially constraining the diversity of perspectives captured within each session.

Fifth, there is a discrepancy between the definition of "family" used, as intended by Goldfarb et al [14], and the sample included. While we advocate for an inclusive definition of

XSL•FO
RenderX

family that includes friends and neighbors, our sample consisted only of familial caregivers (partners, children, and parents). The legal, privacy, and sociotechnical access needs of nonkin caregivers differ from those of immediate family members; yet our findings and features are currently based solely on immediate familial caregivers. Future work should explore whether the requirements of broader social support networks, such as neighbors or friends, align with the insights gathered from these closer familial ties.

Sixth, only one unmet need was identified in the spiritual domain in the results, which is anomalous for cardiac care research, where existential distress is often prominent. The reason for limited results in this domain may be due to patients and relatives perceiving hospitals as institutions focused on physical and psychological care and rehabilitation, rather than existential guidance or meaning. Consequently, they may be less inclined to report spiritual needs in a study focused on remote hospital care and support. Moreover, this may be due to cultural factors in the Dutch sample, where spirituality may have a relatively less prominent role in some subcultures. Future research should determine whether these needs are absent, unexpressed, or possibly addressed outside the formal care system.

Finally, the analysis and construction of the care journey may have been influenced by our individual perspectives and professional interests as medical and design professionals, particularly, our engagement with RPM and digital innovation. Our perspectives may have shaped how certain codes and themes were interpreted and prioritized. We sought to minimize the impact of individual and professional bias by engaging in in-depth consensus discussions and including reflections from transdisciplinary perspectives within the research team.

Conclusion

By exploring the lived experiences of patients and their relatives following MI or peri-OP RPM-supported care trajectories, this study identified experience gaps in RPM-supported cardiac care, especially during preadmission and early postdischarge phases. Through our human-centered design approach, we created an overview of unmet informational, psychoemotional, social, physical, practical, and spiritual needs of families, which informed the generation of RPFM intervention feature ideas.

Our findings demonstrate that the most unmet needs regarding care and support in the RPM-supported care pathways extend beyond hospital admission and discharge. Health behaviors and recovery mostly occur at home, within the patient's relational ecosystem with loved ones. The RPFM intervention feature ideas proposed in this study provide preliminary directions for exploration toward a transition from individually focused monitoring to inclusive, family-centered care and management.

Acknowledgments

We extend our sincere thanks to all the participants for generously sharing their lived experiences, which constitute the core of this paper.

During the writing process of this manuscript, ChatGPT (GPT-5.3; OpenAI) was used to support language editing, including proofreading, suggesting refinements in phrasing, and correcting clarity and grammar. The first author subsequently reviewed and edited the text and takes full responsibility for the final version and its integrity.

Funding

This research was made possible through funding from Medical Delta Healthy Society lifestyle & prevention programme. VRJ has received a FIT-HEART consortium grant (01-001-2024-0621) from the Dutch Heart Foundation.

Data Availability

The depersonalized data supporting this article are available on reasonable request to the corresponding author.

Authors' Contributions

The authors' contributions are presented below following the Contributor Roles Taxonomy [56]:
Conceptualization: JH, VRJ, MK, DEA, VP
Data curation: JH, VRJ (supporting), MK (supporting), DEA (supporting), VP
Formal analysis: JH, VRJ (supporting), MK (supporting), DEA (supporting), VP
Funding acquisition: MK, DEA
Investigation: JH, VRJ (supporting), SMH, MK (supporting), DEA (supporting), VP
Methodology: JH, VRJ, MK, DEA, VP
Project administration: JH, VRJ (supporting), MK (supporting), DEA (supporting), VP
Supervision: VRJ, MK, DEA, VP
Validation: JH (supporting), VRJ, SMH (supporting), NHC, MK, DEA, VP
Visualization: JH, VP
Writing – original draft: JH, VRJ, MK, DEA, VP
Writing – review & editing: JH, VRJ, SMH (supporting), NHC (supporting), MK, DEA, VP

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group overview.
[DOCX File, 18 KB - cardio_v10i1e83055_app1.docx]

Multimedia Appendix 2
The care and well-being journey.
[DOCX File, 170 KB - cardio_v10i1e83055_app2.docx]

Multimedia Appendix 3
Clusters of remote patient and family management interventions.
[XLSX File, 18 KB - cardio_v10i1e83055_app3.xlsx]

References

1. Cowie MR, Bax J, Bruining N, et al. e-Health: a position statement of the European Society of Cardiology. Eur Heart J 2016 Jan 1;37(1):63-66. [doi: 10.1093/eurheartj/ehv416] [Medline: 26303835]
2. Cowie MR, Lam CSP. Remote monitoring and digital health tools in CVD management. Nat Rev Cardiol 2021 Jul;18(7):457-458. [doi: 10.1038/s41569-021-00548-x] [Medline: 33824486]
3. Frederix I, Caiani EG, Dendale P, et al. ESC e-Cardiology Working Group Position Paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
4. Takahashi EA, Schwamm LH, Adeoye OM, et al. An overview of telehealth in the management of cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2022 Dec 20;146(25):e558-e568. [doi: 10.1161/CIR.0000000000001107] [Medline: 36373541]
5. Gannamani R, Castela Forte J, Folkertsma P, et al. A digitally enabled combined lifestyle intervention for weight loss: pilot study in a Dutch general population cohort. JMIR Form Res 2024 Feb 8;8:e38891. [doi: 10.2196/38891] [Medline: 38329792]
6. Hughes A, Shandhi MMH, Master H, Dunn J, Brittain E. Wearable devices in cardiovascular medicine. Circ Res 2023 Mar 3;132(5):652-670. [doi: 10.1161/CIRCRESAHA.122.322389] [Medline: 36862812]
7. Shcherbina A, Hershman SG, Lazzeroni L, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):e344-e352. [doi: 10.1016/S2589-7500(19)30129-3] [Medline: 33323209]
8. Treskes RW, van Winden LAM, van Keulen N, et al. Effect of smartphone-enabled health monitoring devices vs regular follow-up on blood pressure control among patients after myocardial infarction: a randomized clinical trial. JAMA Netw Open 2020 Apr 1;3(4):e202165. [doi: 10.1001/jamanetworkopen.2020.2165] [Medline: 32297946]
9. Artinian NT, Fletcher GF, Mozaffarian D, et al. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010 Jul 27;122(4):406-441. [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
10. Knowles KA, Xun H, Jang S, et al. Clinicians for CARE: a systematic review and meta-analysis of interventions to support caregivers of patients with heart disease. J Am Heart Assoc 2021 Dec 21;10(24):e019706. [doi: 10.1161/JAHA.120.019706] [Medline: 34873919]
11. Lyons KS, Lee CS. The theory of dyadic illness management. J Fam Nurs 2018 Feb;24(1):8-28. [doi: 10.1177/1074840717745669] [Medline: 29353528]
12. Bjørnnes AK, Moons P, Parry M, Halvorsen S, Tønnessen T, Lie I. Experiences of informal caregivers after cardiac surgery: a systematic integrated review of qualitative and quantitative studies. BMJ Open 2019 Nov 11;9(11):e032751. [doi: 10.1136/bmjopen-2019-032751] [Medline: 31719093]
13. Riegel B, Moser DK, Buck HG, et al. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):e006997. [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
14. Goldfarb MJ, Bechtel C, Capers Q 4th, et al. Engaging families in adult cardiovascular care: a scientific statement from the American Heart Association. J Am Heart Assoc 2022 May 17;11(10):e025859. [doi: 10.1161/JAHA.122.025859] [Medline: 35446109]
15. Barriball L, Bremner J, Buchan J, et al. Recruitment and retention of the health workforce in Europe: final report. : European Commission; 2015 URL: https://health.ec.europa.eu/system/files/2016-11/2015_healthworkforce_recruitment_retention_frep_en_0.pdf [accessed 2026-06-11]
16. Rechel B, Grundy E, Robine JM, et al. Ageing in the European Union. Lancet 2013 Apr 13;381(9874):1312-1322. [doi: 10.1016/S0140-6736(12)62087-X] [Medline: 23541057]

XSL•FO
RenderX

17. Houwen J, Karki RS, Janssen VR, Pannunzio V, Atsma DE, Kleinsmann MS. Designing remote patient and family centred interventions: an exploratory approach. Proc Des Soc 2024 May;4:1597-1606. [doi: 10.1017/pds.2024.162]
18. Sanz MF, Acha BV, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16. [doi: 10.5334/ijic.5573] [Medline: 33981193]
19. Bulto LN, Davies E, Kelly J, Hendriks JM. Patient journey mapping: emerging methods for understanding and improving patient experiences of health systems and services. Eur J Cardiovasc Nurs 2024 May 28;23(4):429-433. [doi: 10.1093/eurjcn/zvae012] [Medline: 38306596]
20. Davies EL, Bulto LN, Walsh A, et al. Reporting and conducting patient journey mapping research in healthcare: a scoping review. J Adv Nurs 2023 Jan;79(1):83-100. [doi: 10.1111/jan.15479] [Medline: 36330555]
21. Howard T. Journey mapping: a brief overview. Commun Des Q Rev 2014;2:10-13. [doi: 10.1145/2644448.2644451]
22. Mery CM, Well A, Taylor K, et al. Examining the real-life journey of individuals and families affected by single-ventricle congenital heart disease. J Am Heart Assoc 2023 Mar 7;12(5):e027556. [doi: 10.1161/JAHA.122.027556] [Medline: 36802928]
23. Van Schalkwijk DL, Widdershoven JWMG, Elias-Smale S, et al. ShareHeart: a patient journey map of patients with ischemia and non-obstructive coronary artery disease based on qualitative research. J Clin Nurs 2023 Jul;32(13-14):3434-3444. [doi: 10.1111/jocn.16409] [Medline: 35689371]
24. Kumar SS, Andreakou E, Tzachsan M, Maier AM. Mapping the journeys of atrial fibrillation patients and citizens using wearable devices for remote cardiac monitoring. Proc Des Soc 2023 Jul;3:2575-2584. [doi: 10.1017/pds.2023.258]
25. Rolland JS. Cancer and the family: an integrative model. Cancer 2005 Dec 1;104(11 Suppl):2584-2595. [doi: 10.1002/cncr.21489] [Medline: 16270342]
26. Fitch MI. Supportive care framework. Can Oncol Nurs J 2008;18(1):6-14. [doi: 10.5737/1181912x181614] [Medline: 18512565]
27. Biersteker TE, Boogers MJ, Schalij MJ, et al. Mobile health for cardiovascular risk management after cardiac surgery: results of a sub-analysis of The Box 2.0 study. Eur Heart J Digit Health 2023 Aug;4(4):347-356. [doi: 10.1093/ehjdh/ztad035] [Medline: 37538141]
28. de Frel DL, Schroijen MA, Aardoom JJ, et al. Participatory development of an integrated, ehealth-supported, educational care pathway (diabetes box) for people with type 2 diabetes: development and usability study. JMIR Hum Factors 2024 May 31;11:e45055. [doi: 10.2196/45055] [Medline: 38819880]
29. van Hattem NE, Mijnsbergen NJ, van Os HJA, et al. The effect and implementation of the COVID Box, a remote patient monitoring system for patients with a COVID-19 infection in primary care: a matched cohort study. Fam Pract 2025 Feb 7;42(2):1-7. [doi: 10.1093/fampra/cmae045] [Medline: 39305513]
30. Rakers M, van Hattem N, Hiddink E, et al. Tailoring remote patient management to optimise cardiovascular risk management in primary care: a mixed-methods implementation study informing large-scale implementation. BMC Prim Care 2025 Jul 2;26(1):214. [doi: 10.1186/s12875-025-02906-x] [Medline: 40604485]
31. Treskes RW, van Winden LA, van Keulen N, et al. Using smart technology to improve outcomes in myocardial infarction patients: rationale and design of a protocol for a randomized controlled trial, the box. JMIR Res Protoc 2017 Sep 22;6(9):e186. [doi: 10.2196/resprot.8038] [Medline: 28939546]
32. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field Methods 2017 Feb;29(1):3-22. [doi: 10.1177/1525822X16639015]
34. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
35. Drakenberg A, Sundqvist AS, Fridlund B, Ericsson E. Family involvement in relation to elective open-heart surgery: a critical incident technique study from a family member perspective. Scand J Caring Sci 2025 Mar;39(1):e70003. [doi: 10.1111/scs.70003] [Medline: 39988863]
36. Fleury J, Moore SM. Family-centered care after acute myocardial infarction. J Cardiovasc Nurs 1999 Apr;13(3):73-82. [doi: 10.1097/00005082-199904000-00007] [Medline: 10098007]
37. Tuomisto S, Koivula M, Åstedt-Kurki P, Helminen M. Family involvement in rehabilitation: coronary artery disease-patients' perspectives. J Clin Nurs 2018 Aug;27(15-16):3020-3031. [doi: 10.1111/jocn.14494] [Medline: 29679418]
38. Huriani E. Myocardial infarction patients' learning needs: perceptions of patients, family members and nurses. Int J Nurs Sci 2019 Jul 10;6(3):294-299. [doi: 10.1016/j.ijnss.2019.05.001] [Medline: 31508450]
39. Birtwistle SB, Jones I, Murphy R, Gee I, Watson PM. "Do what you can with a happy heart": a longitudinal study of patient and family members' lived experiences of physical activity post-myocardial infarction. Disabil Rehabil 2022 Jul;44(14):3661-3670. [doi: 10.1080/09638288.2021.1878560] [Medline: 33646893]
40. Blair J, Volpe M, Aggarwal B. Challenges, needs, and experiences of recently hospitalized cardiac patients and their informal caregivers. J Cardiovasc Nurs 2014;29(1):29-37. [doi: 10.1097/JCN.0b013e3182784123] [Medline: 23416934]

XSL•FO
RenderX

41. Davies N. Patients' and carers' perceptions of factors influencing recovery after cardiac surgery. J Adv Nurs 2000 Aug;32(2):318-326. [doi: 10.1046/j.1365-2648.2000.01479.x] [Medline: 10964178]
42. Doering LV, McGuire AW, Rourke D. Recovering from cardiac surgery: what patients want you to know. Am J Crit Care 2002 Jul;11(4):333-343. [Medline: 12102434]
43. Hernández-Palazón J, Fuentes-García D, Falcón-Araña L, et al. Assessment of preoperative anxiety in cardiac surgery patients lacking a history of anxiety: contributing factors and postoperative morbidity. J Cardiothorac Vasc Anesth 2018 Feb;32(1):236-244. [doi: 10.1053/j.jvca.2017.04.044] [Medline: 28803768]
44. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging Patients in Decision-Making and Behavior Change to Promote Prevention. Stud Health Technol Inform 2017;240:284-302. [Medline: 28972524]
45. Schmid ME, Stumm J, Stock S, Girdauskas E. Patient perspectives on health care models in cardiac surgery: a qualitative evaluation. BMC Health Serv Res 2024 Oct 30;24(1):1309. [doi: 10.1186/s12913-024-11791-6] [Medline: 39472923]
46. Brust M, Gebhardt WA, Numans ME, Kiefte-de Jong JC. 'Teachable moments': het juiste moment voor leefstijlverandering [Article in Dutch]. Ned Tijdschr Geneeskd 2020;164:D4835 [FREE Full text]
47. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 3;3(1):e001570. [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
48. Kettunen S, Solovieva S, Laamanen R, Santavirta N. Myocardial infarction, spouses' reactions and their need of support. J Adv Nurs 1999 Aug;30(2):479-488. [doi: 10.1046/j.1365-2648.1999.01103.x] [Medline: 10457251]
49. Pannunzio V, Morales Ornelas HC, Gurung P, et al. Patient and staff experience of remote patient monitoring-what to measure and how: systematic review. J Med Internet Res 2024 Apr 22;26:e48463. [doi: 10.2196/48463] [Medline: 38648090]
50. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576. [doi: 10.1370/afm.1713] [Medline: 25384822]
51. Kaihara T, Scherrenberg M, Intan-Goey V, et al. Efficacy of digital health interventions on depression and anxiety in patients with cardiac disease: a systematic review and meta-analysis. Eur Heart J Digit Health 2022 Sep;3(3):445-454. [doi: 10.1093/ehjdh/ztac037] [Medline: 36712158]
52. Moser DK, Dracup K. Role of spousal anxiety and depression in patients' psychosocial recovery after a cardiac event. Psychosom Med 2004;66(4):527-532. [doi: 10.1097/01.psy.0000130493.80576.0c] [Medline: 15272098]
53. O'Bertos SE, Holmberg D, Shields CA, Matheson LP. Partners' attachment styles and overprotective support as predictors of patient outcomes in cardiac rehabilitation. Rehabil Psychol 2020 Aug;65(3):258-267. [doi: 10.1037/rep0000331] [Medline: 32406738]
54. Zhai S, Chu F, Tan M, Chi NC, Ward T, Yuwen W. Digital health interventions to support family caregivers: an updated systematic review. Digit Health 2023;9:20552076231171967. [doi: 10.1177/20552076231171967] [Medline: 37223775]
55. Zøylner IA, Lomborg K, Christiansen PM, Kirkegaard P. Surgical breast cancer patient pathway: experiences of patients and relatives and their unmet needs. Health Expect 2019 Apr;22(2):262-272. [doi: 10.1111/hex.12869] [Medline: 30636366]
56. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Publ 2015 Apr;28(2):151-155. [doi: 10.1087/20150211]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
CVD: cardiovascular disease
MI: myocardial infarction
peri-OP: perioperative
RPFM: remote patient and family management
RPM: remote patient management

Edited by A Coristine; submitted 27.Aug.2025; peer-reviewed by J Vick, T Rotering; revised version received 21.May.2026; accepted 22.May.2026; published 28.Jul.2026.

Please cite as:
Houwen J, Janssen VR, Hondmann SM, Chavannes NH, Kleinsmann M, Atsma DE, Pannunzio V
Involving Families in Cardiac Care Through Remote Patient and Family Management: Focus Group and Journey Mapping Study
JMIR Cardio 2026;10:e83055
URL: https://cardio.jmir.org/2026/1/e83055
doi:10.2196/83055

© Julian Houwen, Veronica R Janssen, Sara M Hondmann, Niels H Chavannes, Maaike Kleinsmann, Douwe E Atsma, Valeria Pannunzio. Originally published in JMIR Cardio (https://cardio.jmir.org), 28.Jul.2026. This is an open-access article distributed

XSL•FO
RenderX

under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short-Term Arrhythmia Prediction Using AI Based on Daily Data From Implantable Devices: Multicenter Prospective Observational Study

Ignacio Fernández Lozano[1], MD; Joaquín Fernández de la Concha[2], MD; Javier Ramos Maqueda[3,4], MD; Nicasio Pérez Castellano[5], MD; Rafael Salguero Bodes[6], MD; F Javier García-Fernández[7], MD; Juan Benezet Mazuecos[8], MD; Javier Jiménez Candil[9], MD; Tomás Datino[10], MD; Sem Briongos Figuero[11], MD; Javier Paniagua Olmedillas[12], MD; Miguel Nicolás Font de la Fuente[13], MSc; Juan López-Dóriga Costales[13], MSc; Sarai Paz Fernández[13], MBA; Vicente Copoví Lucas[14], MPH

[1]Heart Disease Institute, Hospital Universitario Puerta de Hierro Majadahonda, C. Joaquín Rodrigo, 1, Majadahonda, Madrid, Spain
[10]Heart Disease Institute, Hospital Universitario Quirónsalud Madrid, Madrid, Spain
[11]Heart Disease Institute, Hospital Universitario Infanta Leonor, Madrid, Spain
[12]Heart Disease Institute, Hospital Virgen de la Concha, Zamora, Spain
[13]Monitoring Life S.A, Santa Cruz de Tenerife, Spain
[14]Arrhythmia Network Technology SL, Madrid, Spain
[2]Heart Disease Institute, Hospital Universitario de Badajoz, Badajoz, Spain
[3]Heart Disease Institute, Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain
[4]Heart Disease Institute, Clínica HLA Montpellier, Zaragoza, Spain
[5]Heart Disease Institute, Hospital Clínico San Carlos, Madrid, Spain
[6]Heart Disease Institute, Hospital Universitario 12 de Octubre, Madrid, Spain
[7]Heart Disease Institute, Hospital Universitario de Burgos, Burgos, Spain
[8]Heart Disease Institute, Hospital La Luz, Madrid, Spain
[9]Heart Disease Institute, Complejo Hospitalario de Salamanca, Salamanca, Spain

Corresponding Author:
Ignacio Fernández Lozano, MD
Heart Disease Institute, Hospital Universitario Puerta de Hierro Majadahonda, C. Joaquín Rodrigo, 1, Majadahonda, Madrid, Spain

Abstract

Background: Predictive medicine relies on algorithms to determine clinical treatments tailored to each patient's individual characteristics. Predictive models based on artificial intelligence have shown promise in identifying atrial fibrillation episodes; however, they rarely focus on short-term dynamic prediction.

Objective: This study aimed to evaluate the use of an artificial intelligence model and remote monitoring data extracted from pacemaker devices to predict the onset or worsening of arrhythmias in the short term.

Methods: This was a multicenter prospective observational study in which data from 314 patients were analyzed. A total of 65,243 data sequences were collected, of which 55,532 (85.1%) were used to train the algorithm. This model used 31-day records to predict whether the number of arrhythmic episodes would increase, decrease, or remain the same in the following 14 days.

Results: The sensitivity and specificity of the generated predictions were calculated from 9711 prediction-observation pairs. The global sensitivity was 66.4% (95% CI 64.3%-68.3%), and specificity was 77.4% (95% CI 76.4%-78.4%). For patients with baseline arrhythmia, sensitivity was 76.8% (95% CI 74.6%-78.8%), and specificity was 39.6% (95% CI 35.8%-43.5%). The prediction for patients with no baseline arrhythmia showed a sensitivity of 39% (95% CI 35.1%-43%) and a specificity of 81% (95% CI 80.0%-81.9%). The analysis for the patient subgroup without history of atrial fibrillation (232/314, 73.9%) yielded a 69% sensitivity (95% CI 66.5%-71.5%) and an 80% specificity (95% CI 79.3%-81.3%).

Conclusions: This model was capable of predicting short-term increases or decreases in arrhythmic episodes with reasonable sensitivity and specificity using data collected through remote monitoring of implantable devices. The model's performance is expected to improve progressively as more data samples become available, including demographic data and clinical records.

(JMIR Cardio 2026;10:e85841) doi:10.2196/85841

XSL•FO
RenderX

KEYWORDS

artificial intelligence; AI; atrial fibrillation; AF; machine learning; telemedicine; pacemaker; arrhythmia prediction; predictive medicine

Introduction

In recent years, the use of artificial intelligence (AI) for the diagnosis, prognosis, and evaluation of risk factors and the response to treatment of different pathologies has shown significant advancements. This has had a positive impact on health care systems, improving the efficiency of medical handling and reducing the number of adverse events [1,2].

Cardiac electrophysiology has emerged as one of the fastest-growing subspecialties within cardiology over the last few decades [3,4], which is evidenced by numerous technological advancements aimed at detecting, predicting, and managing cardiac arrhythmias [5,6]. Among those, the prediction of cardiac arrhythmias is especially relevant for patients with atrial fibrillation (AF), who are 5 times more likely to experience thromboembolic events than patients without AF [7].

AI training of information systems based on enormous data quantities from clinical reports and complementary tests from millions of patients not only enables episode prediction in patients diagnosed with AF but also allows for the diagnosis of asymptomatic patients, allowing for early intervention and reducing the number of clinical events [8]. However, cardiac implantable devices can also be leveraged to train AI models for the detection of cardiac arrhythmias. They are particularly well suited for machine learning (ML)–based arrhythmia prediction because they provide continuous, high-fidelity rhythm monitoring over long periods, enabling precise quantification of atrial arrhythmia burden and enhancing diagnostic remote monitoring workflow and device-based therapies [9], with the additional advantage of being already implanted on the patient.

Although several authors have published many AI- and ML-based models for AF prediction in different contexts with good results [10-12], these models were primarily trained and evaluated for predictions over extended periods (often up to 1 year), which provided them with a great amount of input data. However, algorithms supporting short-term, dynamic AF prediction in patients with implantable cardiac devices are scarce. Short-term arrhythmia prediction is uniquely challenging because it requires anticipating nonlinear state transitions driven by heterogeneous and rapidly evolving physiological signals. While long-term prediction is valuable for estimating the risk of AF development in the general population—supporting primary prevention and population-level care planning—dynamic short-term prediction may be particularly impactful in patients who have already been diagnosed with AF. In this context, continuously collected data from implantable cardiac devices enable high-resolution, longitudinal assessment of cardiac rhythm, creating an opportunity for near-term forecasting of AF burden progression. Such device-based prediction could support timely optimization of medical therapy, generation of real-time alerts, and prevention of downstream adverse outcomes. Moreover, in patients with newly diagnosed AF, device-derived dynamic prediction may facilitate closer disease monitoring, guide medication escalation, and enable proactive risk factor modification [10,13].

For that reason, this study aimed to evaluate the use of AI on data extracted from remote pacemaker monitoring for short-term prediction of the onset or increase of AF episodes. Our hypothesis was that data obtained from remote pacemakers could be used to train AI models to predict the onset of atrial arrhythmias with a relatively high sensitivity and specificity without the need for additional datasets.

Methods

Study Design

IA Pacing is a multicenter prospective observational study that started in May 2022 and is ongoing as of January 2026 across 12 centers in Spain. All participating centers used implantable cardiac pacing devices from the same manufacturer (Boston Scientific) with no variation in the telemetry or remote monitoring platform, which was uniformly provided by the vendor. Remote monitoring procedures consisted of daily monitoring.

Ethical Considerations

All procedures performed in this study were conducted in compliance with relevant laws and institutional guidelines and were approved by the ethics committees of the participating centers, including the Comité Ético de Investigación Clínica del Área de Zamora (approved May 25, 2023), Comité Ético de la Investigación Médica of Hospital Universitario 12 de Octubre, Comité Ético de Investigación Clínica de Badajoz (updated November 30, 2021; protocol code: IA-Pacing), Comité Ético de la Investigación Clínica of Hospital Universitario Clínico San Carlos, Comité Ético de la Investigación con Medicamentos of Hospital Universitario Infanta Leonor, Hospital Lozano Blesa (under the protocol approved by the Comité Ético de la Investigación con Medicamentos of Hospital Universitario Puerta de Hierro), Clínica Quirón La Luz and Quirón Pozuelo (both contingent upon favorable review by the Comité Ético de la Investigación con Medicamentos of Hospital Universitario Puerta de Hierro), and Comité Ético de Investigación Médica del Área de Salud de Salamanca (approved March 30, 2022).

Written informed consent was obtained from all participants before enrollment in the study. Participants were informed about the objectives and procedures of the study and their right to withdraw at any time without consequences for their medical care. All procedures performed in this study were conducted in accordance with the ethical principles of the Declaration of Helsinki and complied with applicable European and Spanish regulations governing biomedical research and personal data protection, including the General Data Protection Regulation (Regulation [EU] 2016/679), Ley 14/2007 de Investigación

XSL•FO
RenderX

Biomédica, and Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales.

To protect participant privacy and confidentiality, all data were pseudonymized before analysis. Identifiable information was stored separately from the research dataset in secure institutional systems with restricted access, and only authorized members of the research team were allowed to access the data. Data handling and storage complied with institutional data protection policies and applicable European and Spanish data protection regulations.

Participants did not receive financial compensation for their participation in the study. Inclusion criteria for the study and complete patient recruitment by hospital can be found in Multimedia Appendix 1.

Baseline Characteristics of the Population

Medical records were used to assess patients' baseline characteristics, including demographic data, cardiovascular risk factors, history of arrhythmias, details of the implanted device, procedural data, and pharmacological treatment.

Prediction of Arrhythmia Onset Using AI

To calculate arrhythmia predictions, we used data generated through remote monitoring of patients' pacemakers, which included the following variables:

- General variables—activity level, respiratory rate, and maximum and mean heart rate
- AF-related variables—total number of arrhythmia episodes per day, total duration of those arrhythmias, atrial tachycardia response (mean and maximum ventricular rate during rapid ventricular response events), and apnea-hypopnea index
- Heart failure–related variables—percentage of left ventricular pacing and SD of the average NN intervals every 5 minutes

When considering arrhythmia episodes per day, we took into account only those arrhythmias that were of the "supraventricular tachycardia" or "atrial fibrillation" type as categorized by the device, with no time limit restrictions.

The algorithm divided the data into sequences of 45 days. The first 31 days were used for the algorithm to analyze and predict what would happen in the following 14 days. This time was selected after various tests with different times of training and prediction, with no major changes among them. Furthermore, based on medical criteria, a 14-day predictive time window was considered sufficient to allow for patient scheduling and timely intervention if required. Therefore, after observing the telemetry data for 31 days, a prediction was generated for the following 14 days. The appearance or absence of arrhythmias was divided into 4 categories as follows:

- Increase—the patient already had arrhythmias during the 31 days of observation, and the average number of arrhythmias during the 14 days of prediction increased.
- Decrease—the patient already had arrhythmias during the 31 days of observation, but the average number of arrhythmias during the 14 days of prediction decreased.
- Remains at 0—the patient did not have arrhythmias during the 31 days of observation, and the average number of arrhythmias during the 14 days of prediction remained at 0.
- Goes from 0 to something—the patient did not have arrhythmias during the 31 days of observation, but some arrhythmias occurred during the 14 days of prediction.

Subsequently, to simplify the analysis, the previous 4 categories were regrouped into 2:

- Increased—the average number of arrhythmias during the day increased (categories 1 and 4).
- Maintained or decreased—the average number of arrhythmias during the day was maintained or decreased (categories 2 and 3).

Detailed information about these categories is provided in Multimedia Appendix 2. Similarly, information about data preprocessing and final data volume is provided in Multimedia Appendix 3. Finally, the creation of the predictive model using AI, along with its parameters, its training, and its final results over the *test* data, is detailed in Multimedia Appendix 4.

Furthermore, to generate explanations in the model, a specialized tool called Shapley additive explanations (SHAP) was used, which leverages game theory–based results to identify the variables that had the greatest importance in making each prediction. In Multimedia Appendix 5, the average results for the predictions of each class can be observed, where some clinically relevant variables stand out, such as the presence of apneas or the SD of the average NN intervals. Alongside SHAP, the entropy of the classification was used as a measure of uncertainty to obtain confidence estimations for each prediction, as explained in Multimedia Appendix 6. These tools combined reduce the opacity of AI models, increasing trust in them as a tool for clinical decision-making.

Web Application Development

On the basis of the AI-generated predictions of arrhythmias for the following 14 days and the explanations based on SHAP and confidence estimation, a web application was developed to allow medical personnel to monitor each patient.

Statistical Analysis

Categorical variables are presented as frequencies and percentages, whereas quantitative variables are expressed as means and SDs. The sensitivity and specificity of the arrhythmia prediction were calculated from a 2 × 2 contingency table considering prediction and reality during the follow-up period. Sensitivity and specificity were calculated from 9711 prediction-observation pairs. Statistical analyses were performed using SPSS (version 17.0; IBM Corp).

Results

Data from 314 patients enrolled in the IA Pacing study between May 2022 and May 2024 were analyzed. A total of 65,243 data sequences from implantable devices were collected, of which 55,532 (85.1%) were used to train the algorithm, and the remaining 9711 (14.9%) were used to analyze the sensitivity and specificity of the generated predictions.

XSL•FO
RenderX

Baseline Characteristics of the Population

Table 1 shows the clinical characteristics of the participating patients. A total of 61.3% (192/313) of the patients were male, and the mean age was 77 (SD 9.2) years. The most common indication for pacemaker implantation was atrioventricular block, whereas the predominant pacing site was the right ventricular apex. In total, 81.9% (158/193) of the patients had at least one cardiovascular hospitalization in the year before enrollment.

XSL•FO
RenderX

Table . Baseline characteristics of the patients (N=314). Variables are reported as n/N to account for missing data.

Baseline characteristics	Values
Sex (male), n/N (%)	192/313 (61.3)
Age (y), mean (SD)	77 (9.2)
Pacemaker indication among patients, n/N (%)	
AV[a] block	240/313 (76.7)
Sinus bradycardia	24/313 (7.7)
Other indications	49/313 (15.7)
Pacing type among patients, n/N (%)	
Right ventricular apex	259/310 (83.5)
Left branch	51/310 (16.5)
Type of device used by patients, n/N (%)	
Dual-chamber pacemaker	302/314 (96.2)
Tricameral pacemaker with resynchronizer	12/314 (3.8)
Pacemaker models used by patients, n/N (%)	
Boston Scientific PROPONENT MRI L211	101/314 (32.2)
Boston Scientific ACCOLADE MRI L311	67/314 (21.3)
Boston Scientific PROPONENT EL MRI L231	15/314 (4.8)
Boston Scientific ACCOLADE EL MRI L331	119/314 (37.9)
Boston Scientific VISIONIST X4 CRT-P U225	3/314 (1)
Boston Scientific VISIONIST X4 CRT-P U228	9/314 (2.9)
Patients with CV[b] hospitalizations in the last y, n/N (%)	
0	35/193 (18.1)
1	132/193 (68.4)
>2	26/193 (13.5)
CV risk factors among patients, n (%)	
Arterial hypertension	236/308 (76.6)
Diabetes mellitus	118/296 (39.9)
Dyslipidemia	202/302 (66.9)
Obesity	54/273 (19.8)
Current smoking	22/286 (7.7)
Current alcoholism	14/275 (5.1)
CV history among patients, n/N (%)	
NYHA[c] I	101/187 (54)
Ischemic cardiopathies	52/277 (18.8)
Valvulopathies	77/284 (27.1)
Myocardiopathies	34/279 (12.2)
TIA[d] or stroke	36/277 (13)
Peripheral vascular disease	21/269 (7.8)
AF[e] stroke risk score (CHA_2DS_2-VASc), mean (SD)	3.5 (1.5)
Arrhythmia history among patients, n/N (%)	
AF	42/275 (15.3)
Ventricular arrhythmia	5/313 (1.6)

XSL•FO
RenderX

Baseline characteristics	Values
Other comorbidities among patients, n/N (%)	
COPD[f]	24/273 (8.8)
Asthma	17/270 (6.3)
Pharmacological treatment among patients, n (%)	
Antiaggregants	124/314 (39.5)
Anticoagulants	57/314 (18.2)
ACE[g] inhibitors, ARBs[h], or ARNIs[i]	178/314 (56.7)
β-blockers	79/314 (25.2)
Loop diuretics	66/314 (21.1)

[a]AV: atrioventricular.

[b]CV: cardiovascular.

[c]NYHA: New York Heart Association.

[d]TIA: transient ischemic attack.

[e]AF: atrial fibrillation.

[f]COPD: chronic obstructive pulmonary disease.

[g]ACE: angiotensin-converting enzyme.

[h]ARB: angiotensin II receptor blockers.

[i]ARNI: angiotensin receptor-neprilysin inhibitor.

Regarding cardiovascular risk factors, most patients had hypertension (236/308, 76.6%) and dyslipidemia (202/302, 66.9%), whereas the prevalence of diabetes mellitus was 39.9% (118/296). The mean CHA_2DS_2-VASc score was 3.5 (SD 1.5). A history of AF was present in 15.3% (42/274) of the patients.

Prediction of Arrhythmia Onset Using AI

The sensitivity of the AI-based predictive model was 66.4% for the full sample (1456 true positive cases; 1696 false positive cases; 95% CI 64.3% - 68.3%), whereas specificity was 77.4% (95% CI 76.4% - 78.4%), corresponding to 738 false negative cases and 5821 true negative cases.

Next, we grouped the patients based on whether they had presented arrhythmias during the initial 31-day period used by the algorithm to generate predictions. Two groups were defined: patients with arrhythmias during the baseline period (those classified in the “increase” or “decrease” categories) and patients without arrhythmias during the baseline period (those in the “remains at 0” or “goes from 0 to something” categories).

On the basis of the contingency tables generated, we obtained a sensitivity of 76.8% (95% CI 74.6% - 78.8%) and specificity of 39.6% (95% CI 35.8% - 43.5%) in patients with arrhythmias during the baseline period (1219 true positive cases, 393 false positive cases, 258 true negative cases, and 368 false negative cases) and a sensitivity of 39% (95% CI 35.1% - 43%) and specificity of 81% (95% CI 80% - 81.9%) in patients without arrhythmias during that period (237 true positive cases, 1303 false positive cases, 370 false negative cases, and 5563 true negative cases).

In addition, using data sequences from patients with no history of AF (232/314, 73.9%; 7259 data points), we calculated the model’s sensitivity and specificity for arrhythmia prediction considering both supraventricular tachycardia and AF as the presence of arrhythmia. These were 69% (95% CI 66.5% - 71.5%) and 80% (95% CI 79.3% - 81.3%), respectively.

The raw values from the contingency table for each group can be found in Table 2.

XSL•FO
RenderX

Table . Arrhythmic episode prediction and ground truth for both patient groups.

AI[a] arrhythmia prediction	Follow-up ground truth		Sensitivity (%)	Specificity (%)
	Increased	Maintained or decreased		
Total patients, n			66.4	77.4
Increased	1456	1696		
Maintained or decreased	738	5821		
Patients with arrhythmia in the last 31 days, n			76.8	39.6
Increased	1219	393		
Maintained or decreased	368	258		
Patients without arrhythmia in the last 31 days, n			39.0	81.2
Increased	237	1303		
Maintained or decreased	370	5563		
Patients with no AF[b] history (n=232), n			69.1	80.3
Increased	951	1157		
Maintained or decreased	426	4725		

[a]AI: artificial intelligence.

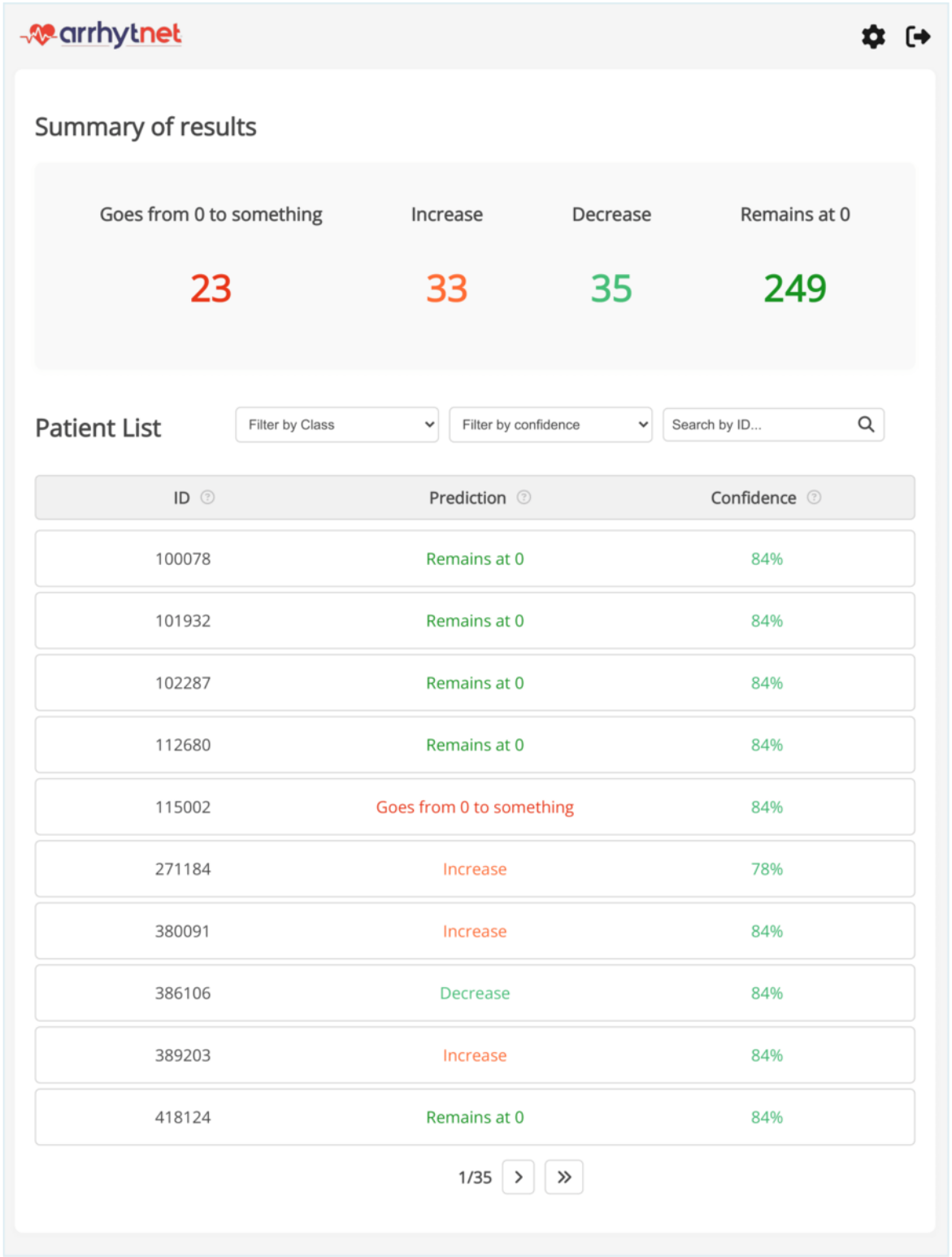
[b]AF: atrial fibrillation.

Web Application Development

The web application interface can be observed in Figure 1 and Figure 2 and presents the prediction along with its confidence score (expressed as a percentage), aiding clinical decision-making. The main view of the web platform shows the list of patients with their predictions and estimated confidence along with a summary of the patients' results and search filters to aid physicians (the number of patients is higher as predictions can be made on patients with 31-45 days of measurement instead of just the >45 days used to validate the AI model). Details of a patient with their prediction, estimated confidence, most important variables for decision-making based on SHAP (pie chart), and variable evolution can also be observed on the web platform.

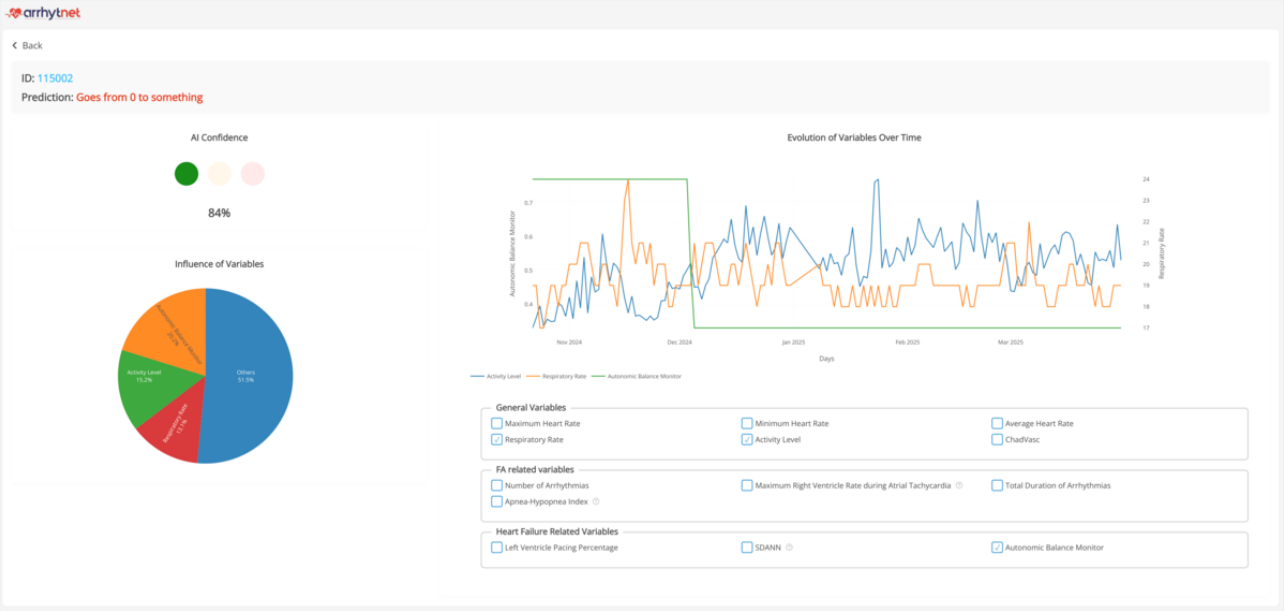
XSL•FO
RenderX

Figure 1. Main view of the web platform showing the list of patients with their predictions and estimated confidence along with a summary of the patients' results and search filters to aid physicians (the number of patients is higher as predictions can be made on patients with 31-45 days of measurement instead of just the >45 days used to validate the artificial intelligence model).



XSL•FO
RenderX

Figure 2. Details of a patient with their prediction, estimated confidence, most important variables for decision-making based on Shapley additive explanations (pie chart), and variable evolution.



Discussion

AI implementation is transforming health care in cardiology, especially in the context of heart rhythm–related disorders. We developed an initial algorithm with information obtained from the remote monitoring of pacemakers that predicted atrial arrhythmia onset within the 14 days following the 31-day monitoring period with a 66.4% sensitivity and 77.4% specificity. This study is, to the best of our knowledge, the first to use data exclusively from remote pacemaker monitoring to dynamically predict arrhythmia onset in patients with atrial arrhythmia in the short term.

Given how important arrhythmia detection is for the identification of patients at risk of having a stroke or systemic embolism, AI algorithms have lately been used for detection, prediction, and treatment tailoring of patients with AF [8]. Focusing on AF episode prediction, several authors have achieved significant advancements in developing AF-predicting algorithms in varying periods from initial measured data. Kao et al [14] developed a model able to predict new arrhythmic episodes in the span of a year with a 98.7% specificity using a database of clinical data gathered for 3 years from older patients with no electrocardiographic data. Similarly, Tiwari et al [15] predicted 6-month AF onset with 75% sensitivity and 85% specificity by developing a model using 200 clinical features from more than 2 million clinical records. Another algorithm worth mentioning developed from clinical records is the one from the Future Innovations in Novel Detection of Atrial Fibrillation study by Nadarajah et al [16] that predicted AF onset within 6 months with a sensitivity and specificity of 78.1% and 73.1%, respectively. Clinical data have also been used to train ML models to detect risk factors for developing AF in asymptomatic patients, as described by Hill et al [10]. Another author used data from 100,000 electrocardiograms (ECGs) taken from 40,000 patients in a mean of 2 years of follow-up to predict the time before a new arrhythmic episode, obtaining similar results to medical record–based predictions [11]. In our study, unlike the studies cited before, no medical record was used to train the system. The algorithm compared daily data from the remote monitoring of an implantable device, similarly to the recently published study by Gregoire et al [13] that trained an ML model using Holter recordings to predict short-term AF episodes. We consider this a strength of our study as only data obtained from the remote monitoring of the implantable device were needed to estimate the risk of experiencing an arrhythmic episode, thus eliminating the need for additional information from medical records.

Although our algorithm does not compute the time before a new arrhythmic episode, it does enable the identification of patients at risk of experiencing episodes of supraventricular arrhythmias in the short term (14 days following the measurement), which would improve their medical care. Our model correctly predicted whether arrhythmias would increase in almost 75% of the cases, with 66.4% global sensitivity and 77.4% specificity. The previously mentioned study from Gregoire et al [13] predicted short-term paroxysmal AF episode onset with higher sensitivity and specificity than ours (86.6% vs 83.0%, respectively). However, it is worth mentioning that, while their approach relies on comprehensive ECG recordings from a Holter monitor, our analysis is based solely on limited, secondary data derived from pacemaker use (eg, daytime heart rate, respiratory rate, or apnea counts). Although these variables provide less information, they may still be highly informative as they are known to be correlated with arrhythmic episodes [17]. Additionally, it is noteworthy that our algorithm was trained using data obtained from a device already implanted in the patient, thereby eliminating the need for additional medical visits. A table comparing the main features of our study to those of the ones mentioned previously can be found in Multimedia Appendix 7.

In patients with no registered history of arrhythmia (232/314, 73.9%), who constitute the group most likely to benefit from early therapeutic interventions, the analysis showed higher values of sensitivity and specificity (69% and 80%,

XSL•FO
RenderX

respectively). Algorithm performance was different according to baseline arrhythmia status (higher sensitivity in patients with arrhythmia in the last 31 days and higher specificity in patients with no episodes in the same period). We think this is due to the different data distribution among these 2 groups as sensitivity and specificity heavily depend on the prevalence of the episode of interest (in this case, arrhythmia).

The fact that our algorithm predicted the occurrence of arrhythmias with a similar performance to that of the model by Tiwari et al [15] (which predicted arrhythmias in a 6-month time window based on clinical records from 2 million younger patients) with relatively few data is a promising starting point to continue enhancing AI training and, therefore, the obtained results. Similarly to Tiwari et al [15], we deem it necessary to optimize AI models to increase their sensitivity and reduce false positives. One strength of our approach is that it requires only a small amount of data for training because the implantable devices continuously monitor cardiac activity. We also used a tool to standardize variable names across devices, which allowed us to combine datasets and increase the amount of data for training, which is another strong point of this study.

It is worth mentioning some limitations of the proposed model. First, the training did not consider demographic variables. Including these factors could increase the predictive capabilities and justify the use of AI over classic risk factor predictors [18]. Moreover, it has been observed that models encompassing more than one type of variable—such as clinical data and ECGs, echocardiograms, or magnetic resonance imaging, for instance—increase prediction sensitivity [19]; this should be taken into account when designing and training future AI algorithms. Second, although the number of prediction and reality sequences was high (more than 9000), data pairs were taken from a cohort of 314 patients, which could be considered low. The size of the cohort could then be increased to augment the training database and refine the output of the model, optimizing AI training and reducing the chance of overfitting the model [20]. It is also worth mentioning that our algorithm did not have external validation as it is a first-phase development, and testing the finalized model on independent data is necessary to draw more specific conclusions on its clinical impact. Additionally, the ability of cardiac devices to differentiate arrhythmia types should be considered when using the extracted data to train the model [21,22]. On this matter, our model did not distinguish among AF, atrial tachycardia, and atrial high-rate episodes, making it hard to fully characterize the clinical relevance of the algorithm. Further research is being conducted considering minimal times, ECG information, and heart rate to fully characterize arrhythmia type. Finally, we extracted the data from devices from the same vendor, and we cannot affirm that this model can be reproducible with other commercial brands. However, fine-tuning methods can be used to train the algorithm using information from other devices. Taking all the aforementioned factors into consideration, further research could focus on increasing the prediction period and on differentiating types of arrhythmias given the clinical implications of de novo episodes of AF and the need for anticoagulation therapy.

Finally, we would like to mention the importance of an intuitive web design to present the obtained data. It has been proven that the use of mobile apps in health care is determined by how user-friendly and intuitive they are and the amount of data that must be input by users (eg, health care professionals) for the apps to function (eg, patient demographics, clinical variables, or diagnostic data) as a higher data entry burden may negatively affect usability and adoption [23]. If an application quickly displays the prediction along with the patient's arrhythmia risk percentage, it is more likely that health care professionals will use it.

In conclusion, this study provides preliminary evidence that changes in arrhythmic burden may be estimated with limited but measurable sensitivity and specificity using data derived from implantable cardiac devices. A graphical overview of the study and its main findings is provided in Multimedia Appendix 8 (Visual Abstract). The predictive performance of the model may improve with further development and validation in larger cohorts that include additional demographic and clinical variables. These elements, together with the user-friendly visualization of predictions in a simple application, suggest that this initial phase of the study may serve as a basis for future research. Further studies are warranted to determine the clinical utility of this approach for the monitoring and management of patients with implantable cardiac pacing devices.

Funding

The authors declare that no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion criteria and patient recruitment.
[DOCX File, 23 KB - cardio_v10i1e85841_app1.docx]

Multimedia Appendix 2
Selection of input and output data.
[DOCX File, 76 KB - cardio_v10i1e85841_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Data volume and preprocessing.
[DOCX File, 21 KB - cardio_v10i1e85841_app3.docx]

Multimedia Appendix 4
Artificial intelligence–based prediction model.
[DOCX File, 80 KB - cardio_v10i1e85841_app4.docx]

Multimedia Appendix 5
Shapley Additive Explanations feature importance.
[DOCX File, 68 KB - cardio_v10i1e85841_app5.docx]

Multimedia Appendix 6
Confidence estimation.
[DOCX File, 97 KB - cardio_v10i1e85841_app6.docx]

Multimedia Appendix 7
Comparison of our model to other predictive models.
[DOCX File, 25 KB - cardio_v10i1e85841_app7.docx]

Multimedia Appendix 8
AI for the Prediction of Atrial Fibrillation or Atrial Tachycardia Episodes in Patients With Pacemakers.
[PNG File, 51 KB - cardio_v10i1e85841_app8.png]

References

1. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: exploring key domains and essential functions. Comput Methods Programs Biomed Update 2024;5:100148. [doi: 10.1016/j.cmpbup.2024.100148]
2. Rizwanullah, Hyderi FA, Obadan GE, Hamad SB, Elessawy H, Ahmadi T. The role of artificial intelligence in predictive medicine: review. J Popul Ther Clin Pharmacol 2024;31(8):1074-1080 [FREE Full text] [doi: 10.53555/jptcp.v31i8.7460]
3. Brugada J. The future of electrophysiology. E-J Cardiol Pract 2022;21 [FREE Full text]
4. Trines SA, Moore P, Burri H, et al. 2024 updated European Heart Rhythm Association core curriculum for physicians and allied professionals: a statement of the European Heart Rhythm Association of the European Society of Cardiology. Europace 2024 Nov 1;26(11):euae243. [doi: 10.1093/europace/euae243] [Medline: 39298665]
5. Isaksen JL, Baumert M, Hermans ANL, Maleckar M, Linz D. Artificial intelligence for the detection, prediction, and management of atrial fibrillation. Herzschrittmacherther Elektrophysiol 2022 Mar;33(1):34-41. [doi: 10.1007/s00399-022-00839-x] [Medline: 35147766]
6. Rahul J, Sharma LD. Advancements in AI for cardiac arrhythmia detection: a comprehensive overview. Comput Sci Rev 2025 May;56:100719. [doi: 10.1016/j.cosrev.2024.100719]
7. Marini C, De Santis F, Sacco S, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. Stroke 2005 Jun;36(6):1115-1119. [doi: 10.1161/01.STR.0000166053.83476.4a] [Medline: 15879330]
8. Karakasis P, Theofilis P, Sagris M, et al. Artificial intelligence in atrial fibrillation: from early detection to precision therapy. J Clin Med 2025 Apr 11;14(8):2627. [doi: 10.3390/jcm14082627] [Medline: 40283456]
9. Antoun I, Alkhayer A, Abdelrazik A, et al. Artificial intelligence and the future of cardiac implantable electronic devices: diagnostics, monitoring, and therapy. J Clin Med 2025 Dec 13;14(24):8824. [doi: 10.3390/jcm14248824] [Medline: 41464726]
10. Hill NR, Ayoubkhani D, McEwan P, et al. Predicting atrial fibrillation in primary care using machine learning. PLoS One 2019 Nov 1;14(11):e0224582. [doi: 10.1371/journal.pone.0224582] [Medline: 31675367]
11. Khurshid S, Friedman S, Reeder C, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. Circulation 2022 Jan 11;145(2):122-133. [doi: 10.1161/CIRCULATIONAHA.121.057480] [Medline: 34743566]
12. Kim IS, Yang PS, Jang E, et al. Long-term PM2.5 exposure and the clinical application of machine learning for predicting incident atrial fibrillation. Sci Rep 2020 Oct 1;10(1):16324. [doi: 10.1038/s41598-020-73537-8] [Medline: 33004983]
13. Grégoire JM, Gilon C, Marelli F, et al. Short-term atrial fibrillation onset prediction using machine learning. Eur Heart J Digit Health 2025 Sep 11;6(6):1159-1168. [doi: 10.1093/ehjdh/ztaf104] [Medline: 41267849]
14. Kao YT, Huang CY, Fang YA, Liu JC, Chang TH. Machine learning-based prediction of atrial fibrillation risk using electronic medical records in older aged patients. Am J Cardiol 2023 Jul 1;198:56-63. [doi: 10.1016/j.amjcard.2023.03.035] [Medline: 37209529]

XSL•FO
RenderX

15. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020 Jan 3;3(1):e1919396. [doi: 10.1001/jamanetworkopen.2019.19396] [Medline: 31951272]
16. Nadarajah R, Wu J, Hogg D, et al. Prediction of short-term atrial fibrillation risk using primary care electronic health records. Heart 2023 Jun 26;109(14):1072-1079. [doi: 10.1136/heartjnl-2022-322076] [Medline: 36759177]
17. Birză MR, Negru AG, Frent M, et al. New insights of cardiac arrhythmias associated with sleep-disordered breathing: from mechanisms to clinical implications-a narrative review. J Clin Med 2025 Mar 12;14(6):1922. [doi: 10.3390/jcm14061922] [Medline: 40142730]
18. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. J Am Heart Assoc 2013 Mar 18;2(2):e000102. [doi: 10.1161/JAHA.112.000102] [Medline: 23537808]
19. Pipilas D, Friedman SF, Khurshid S. The use of artificial intelligence to predict the development of atrial fibrillation. Curr Cardiol Rep 2023 May;25(5):381-389. [doi: 10.1007/s11886-023-01859-w] [Medline: 37000332]
20. Charilaou P, Battat R. Machine learning models and over-fitting considerations. World J Gastroenterol 2022 Feb 7;28(5):605-607. [doi: 10.3748/wjg.v28.i5.605] [Medline: 35316964]
21. Huang JD, Wang J, Ramsey E, Leavey G, Chico TJA, Condell J. Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review. Sensors (Basel) 2022 Oct 20;22(20):8002. [doi: 10.3390/s22208002] [Medline: 36298352]
22. Sperzel J, Hamm CW, Hain A. Over- and undersensing—pitfalls of arrhythmia detection with implantable devices and wearables. Herzschrittmacherther Elektrophysiol 2020 Sep;31(3):273-287. [doi: 10.1007/s00399-020-00710-x] [Medline: 32767089]
23. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]

Abbreviations

AF: atrial fibrillation
AI : artificial intelligence
ECG : electrocardiogram
ML : machine learning
SHAP: Shapley additive explanations

Edited by G Krstačić; submitted 14.Oct.2025; peer-reviewed by H Mautong, JM Santos-Gago; revised version received 30.Jan.2026; accepted 16.Feb.2026; published 18.Mar.2026.

Please cite as:
Fernández Lozano I, Fernández de la Concha J, Ramos Maqueda J, Pérez Castellano N, Salguero Bodes R, García-Fernández FJ, Benezet Mazuecos J, Jiménez Candil J, Datino T, Briongos Figuero S, Paniagua Olmedillas J, Font de la Fuente MN, López-Dóriga Costales J, Paz Fernández S, Copoví Lucas V
Short-Term Arrhythmia Prediction Using AI Based on Daily Data From Implantable Devices: Multicenter Prospective Observational Study
JMIR Cardio 2026;10:e85841
URL: https://cardio.jmir.org/2026/1/e85841
doi:10.2196/85841

© Ignacio Fernández Lozano, Joaquín Fernández de la Concha, Javier Ramos Maqueda, Nicasio Pérez Castellano, Rafael Salguero Bodes, F Javier García-Fernández, Juan Benezet Mazuecos, Javier Jiménez Candil, Tomás Datino, Sem Briongos Figuero, Javier Paniagua Olmedillas, Miguel Nicolás Font de la Fuente, Juan López-Dóriga Costales, Sarai Paz Fernández, Vicente Copoví Lucas. Originally published in JMIR Cardio (https://cardio.jmir.org), 18.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Co-Creation and Usability Testing of a Digital Cardiac Rehabilitation Program (eCardiacRehab): Design and Development Study

Oda Karin Nordfonn[1,2], MSc, PhD; Kristin Johnsen Ramstad[1,3], MSc; Trond Roed Pettersen[1,3], MSc, PhD; Nina Hjertvikrem[1], MA, PhD; Inger-Lise Aamot Aksetøy[4,5], MSc, PhD; Alf Johannes Agcaoili Borge[6]; Eva Gerdts[1,7], MD, PhD, FESC, FAHA; Jan Gunnar Hesthammer[8], MA; Torstein Hole[9,10], MD, PhD; Irene Instenes[1,3], MSc, PhD; Anita Isaksen[1], BSc; Ole Andreas Krumsvik[11], MA; Merethe Landaas[6], BSc; Christian Moltu[12,13], MA, PhD; Svein Rotevatn[1], MD, PhD; Jan Schjøtt[1], MD, PhD; Cathrine Horn Sommersten[14], MSc, PhD; Jan Ove Tryti[15], MD; Tone Merete Norekvål[1,3,7], MSc, PhD, FESC, FAHA

[1]Department of Heart Disease, Haukeland University Hospital, Jonas Lies Vei 65, Bergen, Vestland, Norway

[10]Department of Public Health end Nursing, Norwegian University of Science and Technology, Trondheim, Trøndelag, Norway

[11]Youwell AS, Bergen, Vestland, Norway

[12]Department of Health and Caring Sciences, Western Norway University of Applied Sciences, Førde, Vestland, Norway

[13]Department of Psychiatry, Helse Førde, Førde, Vestland, Norway

[14]Department of Ageing, Health and Care, Kunnskapskommunen Helse Omsorg Vest, Bergen Kommune, Bergen, Vestland, Norway

[15]Sogndal Municipality, Sogndal, Vestland, Norway

[2]Department of Health and Caring Sciences, Western Norway University of Applied Sciences, Stord, Vestland, Norway

[3]Department of Health and Caring Sciences, Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Vestland, Norway

[4]National Advisory Unit on Exercise Training as Medicine for Cardiopulmonary Conditions, St Olav's University Hospital, Trondheim, Trøndelag, Norway

[5]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Trøndelag, Norway

[6]Landsforeningen for hjerte, lunge og hjerneslag, LHL, Bergen, Vestland, Norway

[7]Department of Clinical Science, University of Bergen, Bergen, Vestland, Norway

[8]Section for eHealth, Haukeland University Hospital, Bergen, Vestland, Norway

[9]Medical Department, Helse Møre og Romsdal HF, Ålesund, Møre og Romsdal, Norway

Corresponding Author:
Oda Karin Nordfonn, MSc, PhD
Department of Heart Disease, Haukeland University Hospital, Jonas Lies Vei 65, Bergen, Vestland, Norway

Abstract

Background: Cardiac rehabilitation (CR) is a class IA–recommended treatment and secondary preventive strategy for people with coronary artery disease. eHealth solutions show promise in providing accessible CR. However, few provide a comprehensive and multimodal CR program, and most programs are developed without input from end users.

Objective: The aim of this study was to provide a detailed description of the multidisciplinary co-creation process used to develop a person-centered, multimodal digital CR program that integrated the perspectives of stakeholders and end users throughout, with the rationale to optimize the acceptance and feasibility of the CR program.

Methods: We incorporated a user-centered design with a 3-phase, iterative, multidisciplinary stakeholder co-creation process. This included *defining needs* (phase 1), *development and refinement* (phase 2), and *testing and optimization* (phase 3). Overall, 26 internal stakeholders from the project team participated, including clinicians, researchers, software developers, and user representatives. In addition, 35 external stakeholders (20 clinicians and 15 people with cardiac-related problems) participated in a user-centered development process that included workshops, intervention content development, and usability testing. Data were collected through contextual inquiry, 4 workshops with stakeholders, 12 co-creation sessions with user representatives, 2 usability testing sessions using a think-aloud methodology, standardized usability surveys, and individual and focus group interviews. Intervention content was developed and finalized based on existing evidence, stakeholder input, and user testing. A smoke test and a minimum viable product test were conducted. Data were analyzed using rapid qualitative content analysis and descriptive statistics.

XSL•FO
RenderX

Results: Analyses of data gathered across the 3 phases identified 5 themes important to stakeholders and end users: a program with (1) personalized and tailored solutions with reliable, trustworthy, and evidence-based content (ie, a face-to-face introduction session at the start of the program and a tailored registration page to individually register personal risk factors, goals, and medications); (2) options for feedback (ie, an asynchronous messaging service and tailored face-to-face video consultations with clinicians); (3) peer support (ie, exercise sessions in groups); (4) digital reminders (ie, notifications on a smartphone or tablet); and (5) motivational features (ie, interactive cardiac-specific learning modules on coronary artery disease and secondary prevention). Evidence-based content development resulted in 9 cardiac-specific interactive learning modules, a module for personal health data entry, live exercise sessions, tailored digital follow-up, and an asynchronous messaging service. Readability was assessed using the Gunning Fog Index, with levels ranging from 4 to 13. The overall System Usability Scale score was 86.2 (SD 12.2), indicating excellent usability.

Conclusions: A multidisciplinary stakeholder co-creation process combining well-known evidence-based concepts with stakeholder input identified several areas for improvement and facilitated the development of the digital CR program. Contextually tailored, co-designed interventions may improve relevance and responsiveness to user needs. The feasibility and effectiveness of the program will be further evaluated in the eCardiacRehab trial.

Trial Registration: ClinicalTrials.gov NCT06759805; https://clinicaltrials.gov/study/NCT06759805

(JMIR Cardio 2026;10:e93398) doi:10.2196/93398

KEYWORDS

co-creation; co-design; digital cardiac rehabilitation; multidisciplinary; user experience; eHealth

Introduction

Coronary artery disease (CAD) remains a significant global health burden, accounting for substantial morbidity and mortality despite advances in acute care and secondary prevention strategies [1]. Cardiac rehabilitation (CR) is a well-established, evidence-based secondary prevention intervention shown to reduce mortality, improve functional capacity, and enhance quality of life among people with CAD [2]. Accordingly, CR is a class 1A–recommended intervention in European guidelines [3]. Despite these recommendations, only 55% of countries worldwide offer CR [4]. Although the benefits of CR are well documented as a clinically effective model of care, participation rates range from 14% to 35% [5]. Reasons for low participation in CR include limited access, low referral rates, and low adherence [5]. In Norway alone, more than 11,000 patients undergo percutaneous coronary intervention (PCI) annually, which is an indication for CR. Nevertheless, the national average CR participation rate is only 14% [6].

The increasing integration of eHealth technologies into cardiac care offers new opportunities to address challenges in CR by enabling remote, flexible, and scalable delivery of health care services [7]. Digital CR programs hold promise for improving access and adherence; however, their effectiveness depends critically on user engagement, acceptability, and the alignment of intervention design with user needs and clinical workflows [8,9]. User engagement significantly influences the acceptability of programs. However, optimizing user contributions during program development poses challenges [10].

Co-creation has emerged as a pivotal approach in the development of eHealth interventions, emphasizing the active collaboration of multidisciplinary and diverse stakeholders—patients, clinicians, designers, and researchers—at all stages of the design process [11]. This participatory methodology fosters a deeper understanding of user needs, preferences, and contextual factors, which are essential for creating interventions that are not only clinically effective but also acceptable, feasible, engaging, and motivating for end users [12]. By promoting shared ownership and mutual learning, co-creation enhances the relevance and sustainability of eHealth solutions, ultimately improving implementation, patient adherence, and health outcomes [13]. This methodology has provided beneficial technologies in pain management [14], cancer care [15], and cardiovascular risk reduction [16].

Despite growing recognition of the potential of co-creation, its systematic application in digital CR remains limited, and documentation regarding its impact on program development and implementation is scarce. Addressing this gap is crucial to inform best practices for designing patient-centered digital CR models that effectively translate into real-world settings. Therefore, the aim of this study was to provide a detailed description of the multidisciplinary co-creation process used to develop a person-centered, multimodal digital CR program that integrated the perspectives of stakeholders and end users throughout, with the rationale to optimize the acceptance and feasibility of the CR program.

Methods

Study Context

Despite improved treatment methods, 450,000 Norwegians live with chronic cardiovascular conditions, which not only affect their daily lives but also have implications for clinicians, health care systems, and society at large. The Patient-Reported Outcomes in Cardiology (PROCARD) research group focuses on patient-reported and clinical outcome measures to promote patient-centered care and treatment in cardiology. Through embedded collaboration with regional, national, and international hospitals, municipalities, and academic institutions, the consortium has initiated several large projects aimed at developing and researching equitable CR across the cardiovascular disease (CVD) population. Despite these efforts, many people diagnosed with CVDs are unable to reduce known

XSL•FO
RenderX

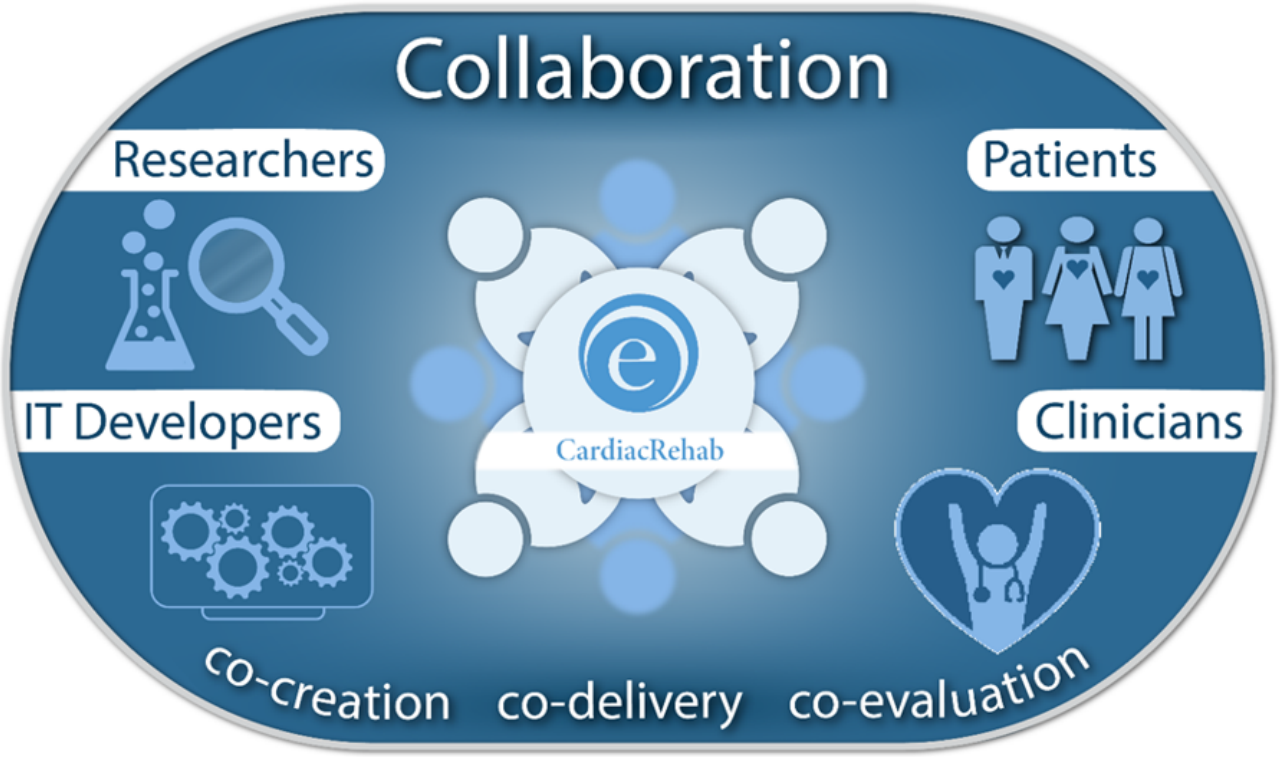
risk factors for CVD. Therefore, PROCARD initiated the eCardiacRehab study embedded in clinical practice. The group has an interdisciplinary approach as the foundation for shared decision-making between patients and clinicians.

Design

Based on a user-centered design [17], the development process (Figure 1) of the digital CR program entailed an iterative 3-staged co-creation process (Figure 2) according to the Medical Research Council (MRC) framework [18]. The project used well-established multidisciplinary CR concepts and class 1A recommendations for secondary prevention of CVD from clinical guidelines [3] as an evidence-based foundation for the co-creation process. The development process was reported in accordance with the GUIDED (Guidance for Reporting Intervention Development Studies in Health Research) checklist (Checklist 1) [19].

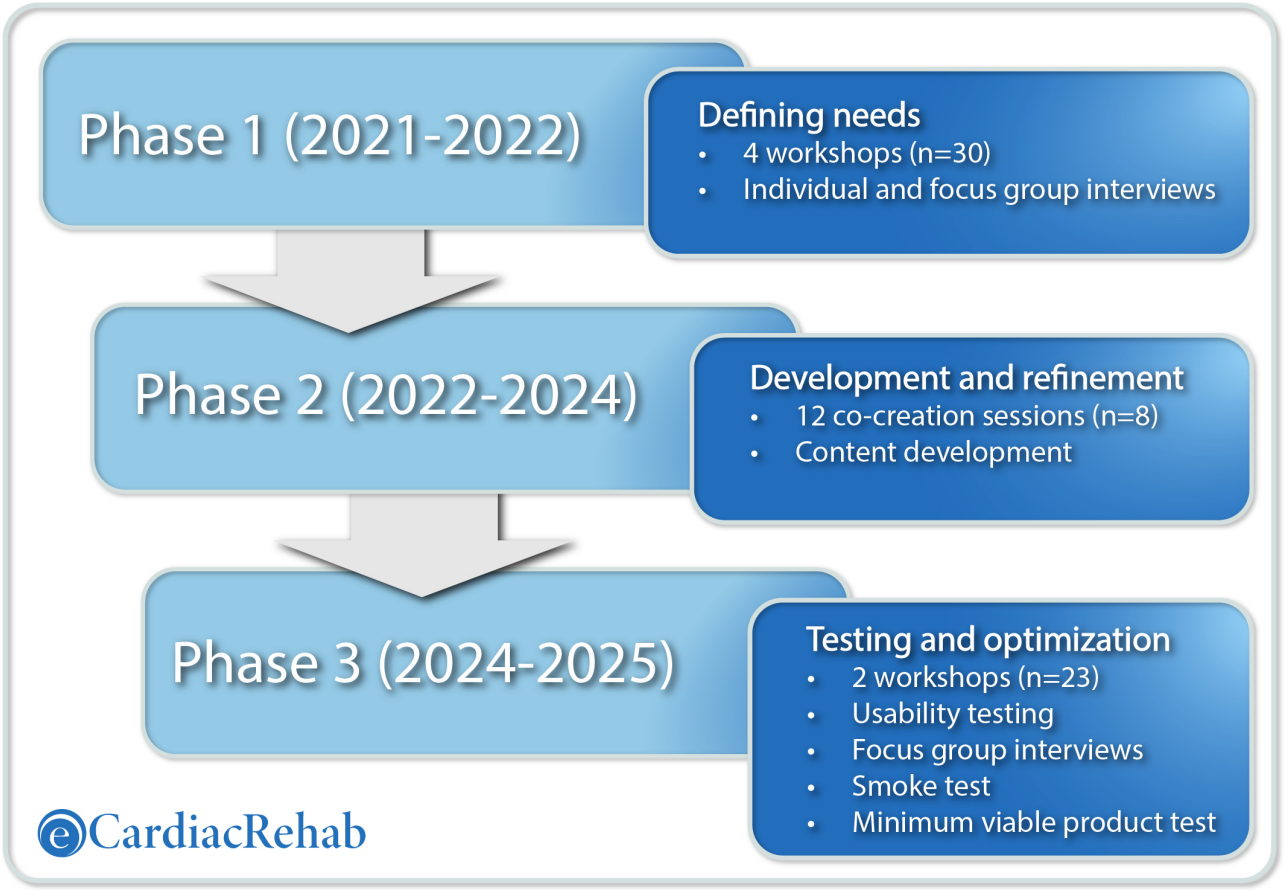
The multidisciplinary project team consisted of researchers, clinicians, a software team of technology developers and designers, and user representatives, who helped ensure that content and materials were presented in an understandable manner. The project team was involved in all phases and met on a regular basis: monthly, weekly, and sometimes more often during the development process from January 2021 until the feasibility phase in January 2025.

Figure 1. Central illustration of the co-creation process in the project.



XSL•FO
RenderX

Figure 2. Overview of the phases in the project's co-creation process.



Recruitment

Health care and research participants were sampled purposively [20] to contribute their competence and experience to the co-creation process, reflecting the multidisciplinary nature of CR (eg, cardiology, nursing, physiotherapy, clinical nutrition, and pharmacology), cardiac and outpatient care (general practitioners [GPs], nurses, and cardiologists), eHealth development, and software and technology.

To be eligible as a user representative, the person had to be aged 18 years or older, have a diagnosis of CVD, and be Norwegian-speaking. A specific goal was to recruit a diverse group of user representatives across age groups and genders through the local branch of the National Association for Heart and Lung Diseases.

Data Collection

The data material was compiled through a 3-phase, iterative process using several data sources (Table 1).

Table . Summary of data sources in the co-creation process of the project.

Phase	Type of data	Which stakeholder
1	• Contextual inquiries • 4 summaries from workshops • 4 transcribed individual interviews	• CONCARD[PCI][a] data (N=3417) [21-24] • Multidisciplinary health care, academic, and IT professionals (n=30) • General practitioners (n=4)
2	• 12 summaries from co-creation sessions	• User representatives (n=8)
3	• 2 reports from usability testing • 3 transcribed focus group interviews • 1 report from smoke test • 1 report from MVP[b] test	• User representatives (n=8, n=15) • User representatives (n=8, n=7, n=8) • Multidisciplinary health care professionals and ICT[c] team

[a]CONCARD[PCI]: Continuity of Care After Percutaneous Coronary Intervention.

[b]MVP: minimum viable product.

[c]ICT: information and communication technology.

XSL•FO
RenderX

Phase 1: Defining Needs

Phase 1 provided insights into end users' needs through multiple methods: *contextual inquiry*, *stakeholder workshops*, and *individual interviews*, ensuring that the project was well anchored.

Contextual Inquiry

The contextual inquiry was based on substantial preparatory work from the CONCARD^{PCI} (Continuity of Care After Percutaneous Coronary Intervention) study, a large Norwegian longitudinal study of people with CAD (N=3417) [21], including a systematic review of modes of eHealth delivery in secondary prevention programs for people with CAD [22], and qualitative and quantitative studies [23,24]. These data provided information about the end users, their needs and requirements for acceptability, and the environment for which the program was designed. Overall, the CONCARD^{PCI} study pointed to the need for new digital CR interventions for people with CAD, including the assessment of health literacy and multimodal health education as important components of eHealth interventions [22,24]. Participants experienced unplanned journeys across care boundaries and did not receive adequate instruction and information on how to integrate health information. They also needed help to schedule clear follow-up appointments, and the study suggested the importance of early follow-up when the patients are motivated. Patients' voices should be taken into greater account when working on follow-up care, pointing to the need for a co-creation approach [23,24].

Workshops With Stakeholders

Clinicians and researchers from collaborating institutions participated in workshops with the eCardiacRehab project team (n=30) facilitated by instructors trained in design thinking from InnoMed. InnoMed is a national competence network for needs-driven innovation in the health care sector, owned and managed by the 4 regional health authorities in Norway. Workshops were conducted digitally during the COVID-19 pandemic (2021 - 2022). Service design methods using personas and patient journeys were used to facilitate user engagement based on existing research recommendations [25,26].

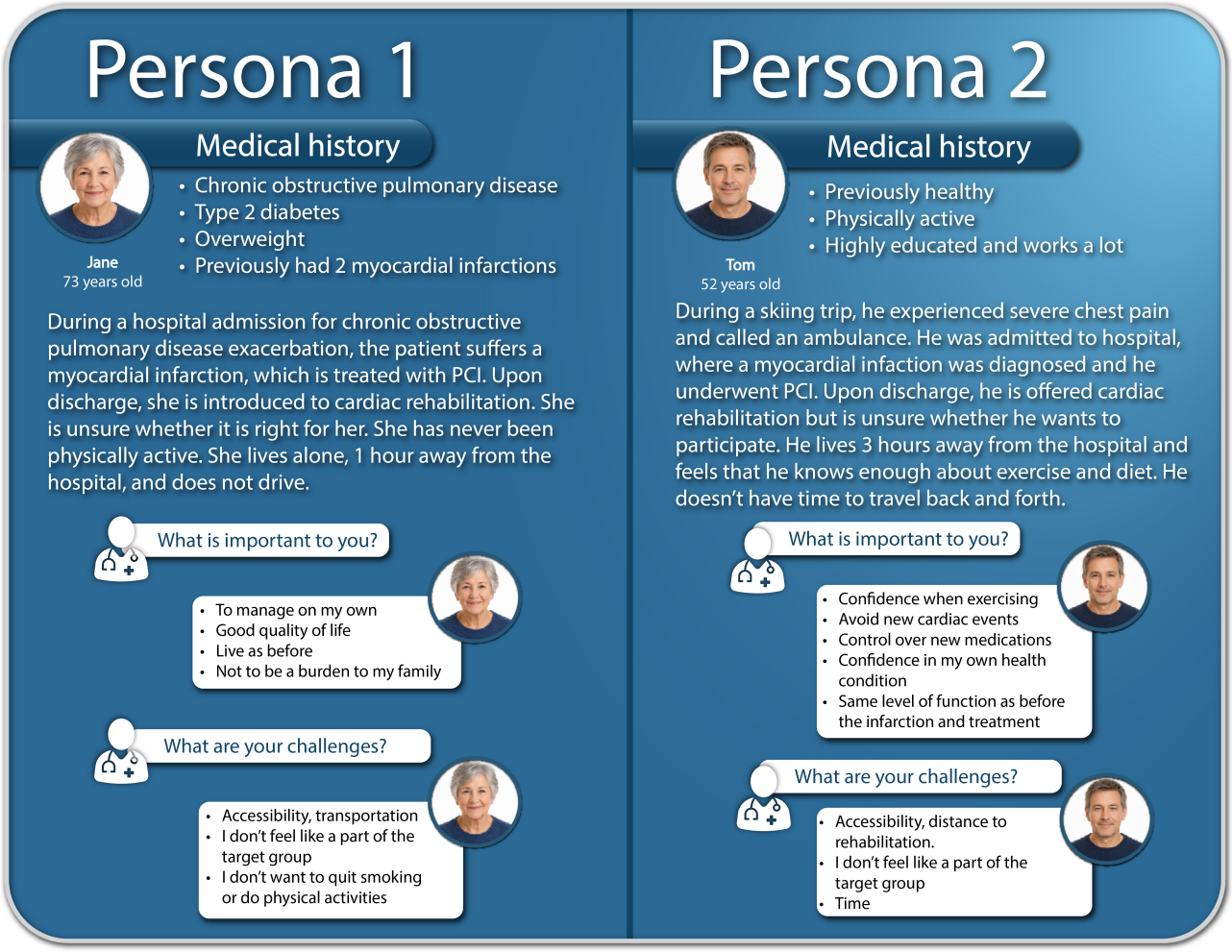
In the workshops, participants were divided into smaller groups of researchers and clinicians from various professional backgrounds. The information gathered in workshop 1 guided the subsequent 3 workshops, which focused specifically on the patient journey and the development process. Workshop 2 focused on recruitment, and workshops 3 and 4 focused on start-up and follow-up in the program. Notes from the workshop discussions were recorded, and materials from group tasks were collected.

Two fictional but representative patient profiles (*personas*; see Figure 3), and a patient journey map (Multimedia Appendix 1) visualizing typical patient pathways were developed [27]. The personas provided a shared understanding of the user profile, created contextual insight into the needs of end users, and helped to keep the focus on the users throughout the process. They represented diversity in CAD risk profiles, background, age, eHealth literacy level, health motivation and frustrations, relationships with clinicians, social support, and other influences on health behavior. The personas included background information (ie, stories to give each persona more depth), coping skills and everyday challenges, an overview of technology skills, and their needs and requirements regarding the eHealth program. The patient journey map described essential components of the rehabilitation pathway and highlighted important elements for digital CR and activities throughout the set period, including contact with clinicians and health care service levels (ie, hospitals and GPs).

Interviews with GPs (n=4) further informed user needs and how GPs could contribute to recruitment and follow-up care in a digital CR program. Data were analyzed using qualitative content analysis [28].

XSL•FO
RenderX

Figure 3. An example of the personas developed in the project. PCI: percutaneous coronary intervention.



Phase 2: Development and Refinement

Phase 2 included an iterative process consisting of 12 co-creation sessions and content development. Dedicated user representatives (n=8), aged 39 to 73 years (Table 2), participated in 12 co-creation sessions from July 2022 to December 2023. All participants lived in an urban area, and each co-creation session lasted approximately 2 hours, including 9 in-person meetings and 3 digital meetings. The purpose was to elicit ideas on design and content features and further explore users' requirements for a digital CR program, with each session informing the next. First, a low-fidelity paper prototype, based on evidence-based topics, clinical guidelines, and logistics from local CR programs, was discussed. Throughout the process, the content was continuously shaped, adjusted, and adapted using an iterative approach based on user representatives' feedback.

All co-creation sessions were facilitated by the second author (KJR) in collaboration with other team members (TMN, NH, TRP, and OKN) and documented in meeting minutes. The first author summarized the material and ensured that it provided essential input into the ongoing development process. The written material was analyzed using qualitative content analysis [28] and sorted into categories representing (1) content, (2) functionality, (3) design, and (4) barriers and facilitators to use (Table 3).

Furthermore, the development of the technology underwent several iterations depending on the topic, with the involvement of the information and communication technology (ICT) team and clinical experts to ensure the use of appropriate language in brief, readily understandable sentences suitable for computer screens. To evaluate the readability of the text, the Gunning Fog Index tool [29] was used. The tool is designed to estimate the years of formal education required to understand a piece of written text on the first reading.

XSL•FO
RenderX

Table . Characteristics of user representatives in co-creation sessions.

Sociodemographic characteristics	User representatives from co-creation sessions (n=8)	New group of user representatives (n=15)
Gender, n (%)		
Women	4 (50)	5 (33.3)
Men	4 (50)	10 (66.7)
Age (y)		
Range	39-73	51-85
Mean (SD)	58.3 (11.2)	69.3 (9.1)
Education, n (%)		
Primary school	—[a]	0 (0)
Upper secondary school	—	5 (33.3)
College or university	—	10 (66.7)

[a]Not available.

XSL•FO
RenderX

Table . Summarized results from 12 co-creation sessions in the project.

Categories	Principles	Comments
Content	• Trustworthiness • Personalization	• Content developed from guideline recommendations for CR[a] and aligned with place-based CR • Information videos from experts in the field • Qualified advice on risk factors, such as tips and menus for alternative diet choices, and exercise training programs on different levels of difficulty • Information in common language and easy to understand • Links to trustworthy external information • Personal health data entry for self-monitoring page for medication, activity, diet, weight, blood pressure, nicotine use, and sleep diary • Goal setting; important for motivation and evaluation • Progression bar to see growth
Functionalities	• Social support • Features	• Dialog by asynchronous message service • Next-of-kin involvement • Group-based activities • Peer support • Easy to navigate, menu function, and program have a logical setup • Self-explaining graphs • Videos and animations on all themes covered • Knowledge quiz in modules to motivate
Design	• Usability	• Notifications, rewards, and reminders • Pleasant color use and animations • Universal design
Barriers and facilitators for use	• Utility	• Proper education in using the program • Motivation in graphs, numbers, and progress bars • Safety in physical test at start-up, contact persons, chat function, real-time exercise sessions • Community, being part of a group • Relatable user stories • Nonstereotypical examples in program

[a]CR: cardiac rehabilitation.

Phase 3: Testing and Optimization

The aim of phase 3 was to test and optimize the user interface and further develop the technical specifications and details of the program.

Usability Testing

The usability testing was conducted from March to April 2024 with user representatives from the co-creation sessions (n=8) and a new group of users (n=15) who were unfamiliar with the program.

The usability testing was conducted face-to-face by facilitators (KJR, TMN, TRP, and OKN) in a data laboratory, where all participants were granted identical computer systems and tools, including headphones (Multimedia Appendix 2). A think-aloud methodology [30] was used to actively engage the participants and elicit continuous feedback, with participants describing their actions and immediate thoughts for each step initiated by the facilitator, both orally and in writing on sticky notes. Observations were summarized following each test and sorted into information groups: (1) usability issues, (2) possible solutions, (3) reported issues, and (4) other input. This approach provided rapid, continuous, yet structured feedback for the development process.

The System Usability Scale (SUS) was used to assess acceptability. SUS measures usability and satisfaction on a scale from 1 (strongly disagree) to 5 (strongly agree). The sum score ranged from 0 to 100, with a cutoff at 68 indicating decent usability, 68 to 80.3 indicating good usability, and greater than 80.3 indicating excellent usability [31]. Data were analyzed descriptively using SPSS Statistics for Windows (version 29.0; IBM Corp).

XSL•FO
RenderX

Focus Group Interviews

After the usability testing, all 23 participants participated in focus group interviews (n=8, n=7, and n=8). The interviews were facilitated by the authors (OKN, TMN, KJR, and TRP) and had a general approach to evaluate the usability testing, the users' experience of the process, and the program content. Data were transcribed verbatim and analyzed at the manifest level using qualitative content analysis [28].

Smoke Test and Minimal Viable Product Test

A smoke test and a minimal viable product (MVP) test were conducted by the research team and representatives from the IT developer, the Section for eHealth at the university hospital, and the Regional Hospital Trust IT (Helse Vest IKT). A smoke test is a preliminary test to ensure that the core functionalities of a software application are working correctly before more extensive testing is performed [32]. Key aspects include ensuring that the application launches successfully, that users can log in and out, that basic navigation within the application works, and that key features, such as search or data entry, are functioning. An MVP test was conducted to reduce risks and ensure a successful launch. It allows developers to gather user feedback early, iterate on the product based on real-world usage, and ultimately build a product that better meets users' needs [33].

Ethical Considerations

The study was reviewed and approved by the Regional Ethics Committee (REK 2015/57 and 459136) and registered at ClinicalTrials.gov (NCT06759805). User involvement was guided by the PRO-Ethics framework [34], a European Union Horizon initiative that emphasizes ethical, inclusive, and transparent participation in research and innovation. This framework supported the design of a participatory process that ensured all participants were treated with respect and had equal opportunities to contribute.

The participants were recruited through a regional patient organization, provided written informed consent, and were informed of the right to withdraw at any time. However, none of the participants withdrew from the study. The majority contributed to all co-creation sessions and participated in workshops, with only a few not participating due to an inconvenient schedule. All contributors were granted full anonymity during the process and in reporting of this study. Participants were involved at all levels of consultation and development, and their voices were given decision-making power in the different steps of the co-creation process, as described.

Results

The 3-phase iterative co-creation process identified several steps of importance for stakeholders and end users, as summarized in Table 4 and detailed briefly in the following sections.

XSL•FO
RenderX

Table . Synthesis of insights from different data sources in the co-creation process.

Phase	Type of data	Insights	Synthesis
1	• Contextual inquiries • 4 summaries/minutes from workshops • 4 transcribed individual interviews	• Bottlenecks and opportunities in cardiac care, from the CONCARD[PCI] study • Insights (1) support to make lifestyle changes, understanding risk factors, and support for service development, (2) practical barriers and facilitators (ie, information, coordination, referral, and inclusion criteria) to recruit, (3) a suggestion for start-up of the program (ie, physical test and consultation with multiple clinicians covering risk factors and motivation), content in the digital CR[a] program, and estimated costs and potential for funding	• Start-up consultation • Planned patient journey • Goal setting • Final consultation
2	• 12 summaries/reports from co-creation session	• Insights into (1) content, (2) functionality, (3) design, and (4) barriers and facilitators to use	• Tailored solution—start-up page with menu and personal health data entry for self-monitoring page • Options for feedback—asynchronous message service • Peer support—live workout sessions led by a physiotherapist and live medical yoga sessions • Digital reminders—options for notifications and reminders • Motivational features—9 interactive cardiac-specific learning modules
3	• 2 reports from usability testing • 3 transcribed focus group interviews • 1 report from smoke test • 1 report from MVP[b] test	• Insights to (1) usability issues, (2) possible solutions, (3) reported issues, and (4) other input • Insights highlighted the importance of motivation, through (1) personal contact throughout, and in the program, (2) to continue the behavioral change (eg, exercise and diet), linked to easy access, good information on the right level, motivational features, and a focus on "every step and activity is better than the sofa," (3) important for the participants to feel safe, that they were assessed by qualified clinicians prior to startup of the program and that they were not on their own. The reach of the program and program fit for different participant groups (ie, women and older adults) were emphasized.	• Inclusive of phase 2 synthesis

[a]CR: cardiac rehabilitation.

[b]MVP: minimal viable product.

XSL•FO
RenderX

Phase 1: Defining Needs

Workshops With Stakeholders

The initial workshops with stakeholders (n=30) contributed to anchoring the project in the different system levels of health care service. From the developed personas and patient journey, participants identified several ways that a digital CR program could meet users' needs, including support for making lifestyle changes, understanding CAD risk factors, and support for service development. Core elements, including practical barriers and facilitators (ie, information, coordination, referral, and inclusion criteria), to recruitment were discussed. The workshops provided feedback on the start-up of the program (ie, physical testing and consultation with multiple clinicians covering risk factors and motivation), the content of the digital CR program, and suggestions regarding the estimated costs and financial opportunities in the health care system.

Interviews With GPs

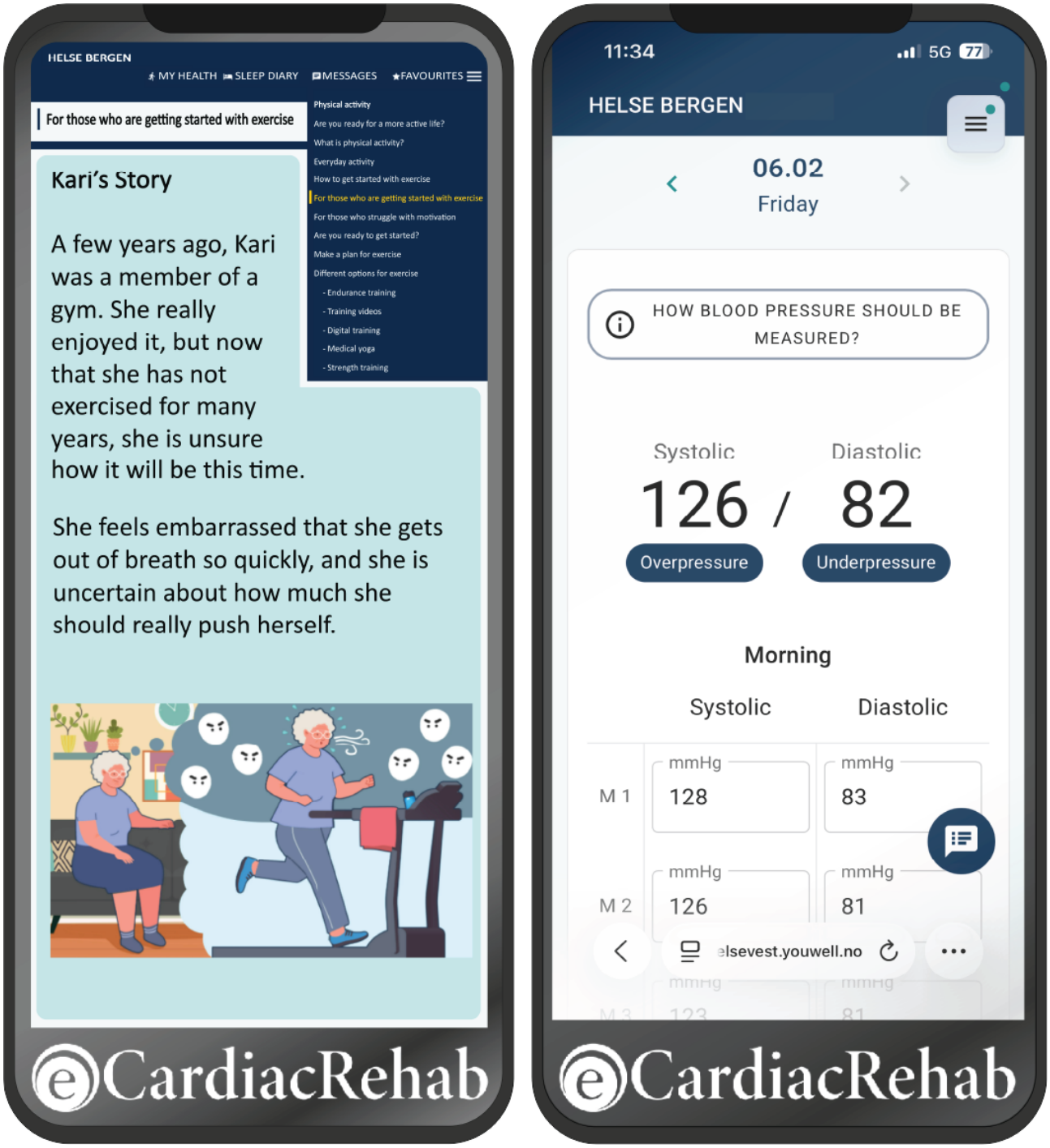
GPs expressed that a digital solution would be suitable for reaching more patients and offering a comprehensive service that could uphold patients' daily lives, as they could continue working during participation in the program. Furthermore, GPs saw themselves as motivators for healthy lifestyle changes through a digital CR program, actively referring people to CR, and viewed the digital CR program as a learning tool to better enhance patients' resources.

Phase 2: Development and Refinement

The 12 co-creation sessions contributed to the development of the prototype from low to high fidelity. The target audience representation was adequate, as the participants included both women and men within the target age range who had CVD, and all used a mobile phone and/or computer in daily life. Prototypes of increasing fidelity were developed, with testing of the content, animations, videos, language, navigation, and labels, including appealing features such as the personal health data entry for self-monitoring (Figure 4).

XSL•FO
RenderX

Figure 4. Example of cardiac rehabilitation trajectory and data entry for self-monitoring module developed in the process.



Throughout the co-creation sessions, a range of viewpoints and preferences regarding the use of a digital CR program were revealed. The role of technology in day-to-day life varied in the group. While most users foresaw great benefits from a digital CR solution, some also highlighted uncertainty about its benefits. Comments made by the user group illustrate the link between categories and end-product features representing (1) content, (2) functionality, (3) design, and (4) barriers and facilitators to use (Table 3). At the start of the co-creation sessions, users listed their preferred attributes for the digital program, highlighting the following: a physical consultation at the start of the program, a chat function during the rehabilitation period, counseling on diet, the involvement of next-of-kin, goal-setting, and progression measures. During the co-creation sessions, users had access to high-fidelity prototypes and specific content, such as modules, animations, and videos. An example of feedback was the setup of the modules, where the module on psychosocial health was deemed so important that it was prioritized as the third component in the sequence. Another example was the suggestion to develop a module on female heart health and a resource library to gather trustworthy information. Although the number of participants in the group was small, a saturation effect in terms of the feedback and input from participants was reached.

Building on the feedback and discussions within the project team, a prototype was developed that included a start page with

XSL•FO
RenderX

a menu and 9 cardiac-specific interactive learning modules in 5 main components: (1) physical activity and exercise, (2) healthy lifestyle education, (3) psychosocial health, (4) self-monitoring and self-management, and (5) medication adherence. Furthermore, a registration module for health data entry for self-monitoring (medications, weight, blood pressure, blood sugar, activity, diet, sleep, mental health, and nicotine status) was designed (Figure 4). The program also includes real-time guided workout sessions led by physiotherapists, real-time medical yoga sessions, and educational videos in all modules. An asynchronous messaging service and scheduled digital consultations ensure communication with the multidisciplinary CR team throughout the program. Moreover, 2 user stories were developed to illustrate different CR trajectories (Figure 4), including variations in acute care pathways across the program's modules. These user stories were included in the program to provide patients with relatable scenarios, compensating for the lack of peer interaction typically present in center-based CR. The readability screening revealed a suitable Gunning Fog Index, with optimal readability levels identified as falling within the range of 4 to 13. This indicates that the text is comprehensible to readers with 9 or fewer years of formal education. The overall SUS score was 86.2 (SD 12.2), indicating excellent usability.

Phase 3: Testing and Optimization

Interviews With Focus Group

The focus group interviews following usability testing provided rich conversations and further informed the program development. Participants highlighted the importance of motivation, both through personal contact throughout the program and within the program. Motivation to continue the behavioral change (ie, exercise and diet) was linked to easy access, information at the right level, motivational features, and a focus on "every step counts, and activity is better than the sofa." To be able to participate in a digital CR program, it was important for the participants to feel safe; they wanted reassurance that an elevated heart rate was not worrisome, that they could exercise safely at their own pace, that they were assessed by qualified clinicians prior to the start of the program, and that they were not on their own. The reach of diverse groups with a program that fit all users was emphasized.

The second and third focus group interviews were conducted with participants (n=7 and n=8) unfamiliar with the program. Several had previously participated in center-based CR and stated that if the digital solution had been available to them at that time, they would have preferred digital CR over center-based CR. They expressed a need for trustworthy and validated content and reported that searching the internet often was a suboptimal solution; hence, this program filled an information gap. The program was considered user-friendly, with a good overview and easy to access and use. One participant said:

> *I also think it's fun to be able to help develop something that others will need later, and something that I was looking for myself during the first months before I started CR, because there was little information then. I was Googling. It's nice to have everything in one place in a programme like this.* [P7]

Some had concerns regarding the lack of fellowship with peers during the CR period and expressed that this would be absent in a digital solution, in both educational and training sessions. Still, there were those who preferred home-based group exercise rather than having to travel a long distance. They also appreciated the granular information level and asked for digital access to the program for next-of-kin. They saw themselves as curious and interested in digital solutions.

Smoke Test and MVP Test

In total, 33 defects were detected during the smoke test. Of these, 4 were categorized as "B—Major" and 29 as "C—Minor" according to the application life cycle management (ALM) system. The major defects included a lack of an alert when participants repeatedly registered elevated blood pressure, the option to register medication consumption for a previous period, malfunctioning links to external web pages, and the inability of clinicians to tailor the program according to patients' needs (ie, to remove the nicotine module if the patient was a nonsmoker). Minor defects were related to data entry for self-management, visualization of data entry, missing text in educational modules, typographical errors, and registrations not being shown on the clinician's dashboard.

Furthermore, at the time of testing, the program was delivered as a responsive web application rather than a native application. Therefore, display and navigation flaws were detected when the content was viewed on smaller mobile phone screens. In order to avoid these challenges, eCardiacRehab was made available both as a web and mobile application (hybrid app). The mobile application can be used on both Android and iOS platforms. Furthermore, the possibility of using biometrics when logging in was established.

In total, 108 defects were detected during the MVP test. Of these, 6 were categorized as "B—Major," 99 as "C—Minor," and 3 as "D—Observation" according to the ALM system. The major defects were related to a lack of contrast in the visualizations of patients' health data entry, and thereby not adhering to the legal requirements for universal design of ICT in Norway.

Privacy by Design, Software Flow, and Technical Architecture

The CR program is developed in accordance with the principles of Privacy by Design. This entails that the system is designed and maintained with data protection as a core requirement. Measures include security testing, secure coding practices, quality standards for design and implementation, and robust procedures for the operation and governance of the information system. Regular penetration testing is performed to identify and mitigate potential vulnerabilities.

The design and development were in accordance with the European General Data Protection Regulation of 2018 [35]. Stringent access controls and high-security requirements (level 4) for authentication were implemented, in addition to safeguards ensuring adherence to fundamental data protection

XSL•FO
RenderX

principles. Access to personal data is strictly limited to the patient's clinician, the system administrator, and designated superusers within the hospital. The digital tool operates on a one-to-one basis, meaning that each patient is linked to their respective clinician. Comprehensive logging functionality ensures full traceability. The tool also provides an overview of assigned access rights. A formal log review procedure to be carried out by the system administrator has been established.

The cloud solution Microsoft Azure is used as a subcontractor, with a data center located in Norway, which is risk-assessed by the regional hospital trust's ICT department. The solution is hosted in Azure using platform as a service (PaaS). App services are used for APIs and websites, and Microsoft SQL Database is used for databases. All data transmission is encrypted over HTTPS and complies with the requirements of the Norwegian National Security Authority and Section 2.2 of the NORMEN (Norwegian Code of Conduct for Information Security and Data Protection in the Health Care Sector) guidelines.

The database (MS SQL) is encrypted with transparent data encryption. This provides "encryption at rest," which prevents data from being read in the event of a data breach or a similar incident that provides direct disk access to the database files.

A risk assessment analysis of the platform (ROS-ID 965) and the use of activity trackers (ROS-ID 25043) was performed. Furthermore, a data protection impact assessment was conducted (2025/2926 - 4).

Discussion

Principal Findings

In this study, we describe the iterative co-creation process and design of eCardiacRehab, a multimodal digital CR program. To our knowledge, this is one of the few studies that combine thorough stakeholder and user involvement and research their preferences. Moreover, this approach combined stakeholder and user perspectives with evidence-based knowledge to develop a digital home-based CR program that covers all guideline-directed components of CR.

The development process combined evidence-based and user-centered approaches, a recommended but underused approach to the development of eHealth solutions [36]. In this study, we incorporated all guideline-recommended modalities of CR. Although some initiatives, such as CoroPrevention [37] and INTERCEPT [38], have recently taken the multimodal approach forward, prior digital interventions have focused primarily on physical activity or exercise training and have lacked the full spectrum of core components of CR, such as lipid and diabetes management, nutrition, psychosocial support, and smoking cessation [39]. Given the importance of contextual factors in shaping intervention uptake and effectiveness, interventions should be adapted to local, regional, and national circumstances. Co-designing interventions within users' own context can help ensure that the intervention is aligned with user needs, priorities, and implementation conditions [10,36]. Our iterative co-creation process identified 5 themes of importance for stakeholders in meeting the end users' needs. These themes included a tailored solution, options for feedback, peer support, digital reminders, and motivational features. Areas for improvement regarding the 5 themes were identified through contextual inquiries, stakeholder workshops, and iterative content development based on evidence-based strategies for CR. In general, ways to optimize the co-creation process from a user perspective are missing [10], and our results suggest avenues for shaping the process through the use of moderators, guidance, and continuous engagement.

Patient Involvement

This study highlights the importance of thorough patient involvement in co-creation sessions over an extended period. Such engagement is not merely beneficial but essential for understanding and integrating the needs and contexts of intended users [40], emphasizing how critical it is to prioritize user perspectives in the development of effective eHealth solutions [41,42]. A recent systematic review [41] identified usability and user-friendliness as prominent technology-related factors for health care professionals' adoption of digital health interventions. This underscores the value of co-design methodologies in developing interventions that are both acceptable and implementable in clinical practice. Continuous patient involvement is essential not only for maintaining engagement but also for building trust and ensuring that participants feel safe and valued throughout the process [42]. It enabled users to address the core needs for a tailored solution, including options for feedback, peer support, digital reminders, and motivational features. Adopting a tailored approach is crucial for enhancing digital CR programs [43]. Engaging patients throughout the co-creation trajectory enabled the collection of essential feedback that informed the design decisions. The MRC framework [18] emphasizes that meaningful engagement of patients can maximize the potential of novel interventions that are likely to positively affect health. Empirical evidence substantiates that active user participation in development enhances the likelihood of successful adoption and seamless integration into users' daily lives [11,44]. Furthermore, user involvement serves to diminish barriers to technology adoption, fostering a sense of ownership and agency among users, both of which are crucial determinants of the sustainability and effectiveness of a digital CR program [45].

Principal Methodology

Patient and public involvement is essential when developing eHealth interventions [18]. We initiated the co-creation process involving stakeholders in workshops using personas, which are fictional representations of characteristic real users [27], to better target the end users' needs. The deployment of personas can function as an instrumental mechanism for acquiring a nuanced understanding of the user experience, fostering empathy, and elucidating user needs [46]. We constructed robust personas that helped focus the stakeholder workshops on the actual needs and goals of patients. A co-creation process is time-consuming due to the necessity for multiple feedback cycles, and it requires significant resources, both financial and human [11]. The use of personas may contribute to less extensive processes and contribute to the enhancement of self-management behaviors in digital CR [47]. This program was developed to engage a broad range of users, promote

XSL•FO
RenderX

diversity, and address inequities in eHealth solutions [48]. Personas may elucidate how diverse patient demographics—such as women, older adults, or individuals with multiple comorbidities (persona 1)—exhibit heterogeneous requirements that necessitate tailored CR programs. Personas may also help narrow the design options, thereby enhancing the effectiveness of co-creation processes [36]. Using personas may also facilitate deeper user engagement, thereby advancing the overall user experience when developing a digital CR program [27].

Strengths, Limitations, and Future Directions

The active user participation throughout the design process, with a diverse and representative group, is a strength that can lead to increased adoption and advocacy of the program. The same principal investigator (PI) and team followed the process and supervised all development phases, ensuring continuity in all aspects. The PRO-Ethics framework [34] was essential in enabling meaningful engagement and ensuring that participants' perspectives were genuinely integrated into the development process.

However, this process was not without challenges. Despite strong user involvement, users may lack the technical expertise to provide insightful feedback on all aspects of the design. Therefore, while co-creation holds substantial promise for driving innovation and fostering user-centered design, it necessitates careful management to balance its benefits against inherent challenges.

One key issue is that eHealth approaches are not finalized after a pilot or main trial, but need a strategy for co-creation and revision across the full life cycle. As co-creation processes continue to evolve, there are several future directions to enhance their effectiveness and impact on health service development. One promising avenue is the integration of advanced technologies, such as AI and machine learning, into co-creation platforms in health care [49]. These technologies have the potential to analyze user data more effectively, providing objective insights that can help tailor the co-creation experience by identifying patterns beyond and across individual user preferences and behaviors.

Conclusions

This study offers insights into how to take a user-centered approach to the design and development of an evidence-based digital CR program to secure a tailored solution. The study involved stakeholders and users in all the phases of the iterative process, pointing out important steps for developing useful and meaningful interventions for patients using stakeholder workshops and co-creation sessions. In addition, to inform the development process, this study provides a practical example for future studies. Co-creation sessions with patient engagement over an extended period and the use of personas are recommended. To test usability, feasibility, and acceptability, as well as the potential efficacy of the program, further research is needed. A feasibility study is currently underway to optimize eCardiacRehab in preparation for a full-scale randomized controlled trial.

Acknowledgments

We acknowledge a diverse group of experts and users who have contributed significantly to the co-creation process: Hege Andersen, Helge Anthoniussen, Stine Austgulen, Christ Berge, Anne Grethe Danielsen, Irene Drotningsvik, Marieie, Hayes, Ingvild Hernar, Stig Igland, Marjolein Iversen, Arnstein Ivesdal, Marianne Jensen, Morten Kolltveit, Ingeborg Kløve-Graue, Nils Medaas, Ingunn Mjøs, Ragnhild Sekse, Marit Solheim, Elfrid Herre Staveland, Irene Valaker, and Eirik Østvold.

We are also thankful to our international Scientific Advisory Board, chaired by Professor Heather Allore (Yale University), Professor Maria Bäck (Linköping University), Professor Christi Deaton (University of Cambridge), Professor Robyn Gallagher (University of Sydney), Professor Heather Hadjoustrapolous (University of Regina), and Professor Ann-Dorthe Zwisler (University of Copenhagen). We thank Marie Hayes for assistance with the figures. We also thank the full group of eCardiacRehab Investigators.

Generative AI was not used in any portion of the manuscript generation.

Funding

eCardiacRehab received a process evaluation grant from InnoMed, funded by all Norwegian Regional Health Authorities; an innovation grant from the Western Norway Regional Health Authority (R-10930), a research grant from the Western Norway Regional Health Authority (F-12624), and a major research grant from the Norwegian Research Council (ES726489). TRP is funded by a postdoctoral fellowship from the Western Norway Regional Health Authority (912295).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Expert and steering group member: SR
Key project member: NH, II
Local principal investigator and steering group member: TH
Mini team leader for digital development: OAK

XSL•FO
RenderX

Mini team leader for physical activity: ILAA
Mini team leader for primary care involvement and collaboration: CHS
Mini team leader for referral and inclusion: OKN
Mini team leader for screening and patient-centered goals: AI
Mini team leader for technological support: JGH
Mini team leader for user involvement: KJR
Patient representative: AJAB, ML
Principal investigator: TMN
Project coordinator: TRP
Steering group member: JOT
Study design: All authors
Work package leader for adherence to secondary prevention strategies: JS
Work package leader for mental health: CM
Work package leader for women: EG
Writing - original draft: OKN
Writing - review & editing: All authors
All authors read and approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of a journey map developed in the eCardiacRehab.
[PNG File, 5137 KB - cardio_v10i1e93398_app1.png]

Multimedia Appendix 2
Image 1: picture from usability testing laboratory.
[PNG File, 296 KB - cardio_v10i1e93398_app2.png]

Checklist 1
GUIDED checklist.
[PDF File, 204 KB - cardio_v10i1e93398_app3.pdf]

References

1. Lindstrom M, DeCleene N, Dorsey H, et al. Global burden of cardiovascular diseases and risks collaboration, 1990-2021. J Am Coll Cardiol 2022 Dec 20;80(25):2372-2425. [doi: 10.1016/j.jacc.2022.11.001] [Medline: 36517116]
2. Dibben GO, Faulkner J, Oldridge N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 7;44(6):452-469. [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
3. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). Rev Esp Cardiol (Engl Ed) 2022 May;75(5):429. [doi: 10.1016/j.rec.2022.04.003] [Medline: 35525570]
4. Supervia M, Turk-Adawi K, Lopez-Jimenez F, et al. Nature of cardiac rehabilitation around the globe. EClinicalMedicine 2019;13:46-56. [doi: 10.1016/j.eclinm.2019.06.006] [Medline: 31517262]
5. Li L, Ringeval M, Wagner G, Paré G, Ozemek C, Kitsiou S. Effectiveness of home-based cardiac rehabilitation interventions delivered via mHealth technologies: a systematic review and meta-analysis. Lancet Digit Health 2025 Apr;7(4):e238-e254. [doi: 10.1016/j.landig.2025.01.011] [Medline: 40023729]
6. Norekvål TM, Bale M, Bedane HK, Hole T, Ingul CB, Munkhaugen J. Cardiac rehabilitation participation within 6 months of discharge in 37 136 myocardial infarction survivors: a nationwide registry study. Eur J Prev Cardiol 2024 Nov 18;31(16):1977-1980. [doi: 10.1093/eurjpc/zwad350] [Medline: 37943676]
7. Tromp J, Jindal D, Redfern J, et al. World Heart Federation roadmap for digital health in cardiology. Glob Heart 2022;17(1):61. [doi: 10.5334/gh.1141] [Medline: 36051317]
8. Chaturvedi A, Prabhakaran D. Transforming cardiovascular care with digital health: the past, progress, and promise. JACC Adv 2024;3(9):101183. [doi: 10.1016/j.jacadv.2024.101183] [Medline: 39220713]
9. Kumar D, Hazra PK, Gharat CH. Long-term outcomes of digital cardiac rehabilitation. JACC Asia 2025 Oct;5(10):1225-1238. [doi: 10.1016/j.jacasi.2025.08.007] [Medline: 41067820]

XSL•FO
RenderX

10. Tremblay M, Hamel C, Viau-Guay A, Giroux D. User experience of the co-design research approach in eHealth: activity analysis with the course-of-action framework. JMIR Hum Factors 2022 Aug 9;9(3):e35577. [doi: 10.2196/35577] [Medline: 35943783]
11. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health—a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2):3222211. [doi: 10.17061/phrp3222211] [Medline: 35702744]
12. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
13. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. Milbank Q 2016 Jun;94(2):392-429. [doi: 10.1111/1468-0009.12197] [Medline: 27265562]
14. Ledel Solem IK, Varsi C, Eide H, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020 Jan 21;22(1):e15889. [doi: 10.2196/15889] [Medline: 31961331]
15. Difrancesco S, Bauert MM, Lehmann C, et al. Collaborative design and development of a patient-centered digital health app for supportive cancer care: participatory study. JMIR Hum Factors 2025 Nov 11;12:e73829. [doi: 10.2196/73829] [Medline: 41217811]
16. Neubeck L, Coorey G, Peiris D, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) web application. Int J Med Inform 2016 Dec;96:24-37. [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]
17. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge WS, editor. Encyclopedia of Human-Computer Interaction: Sage Publications; 2004:445-456.
18. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
19. Duncan E, O'Cathain A, Rousseau N, et al. Guidance for Reporting Intervention Development Studies in Health Research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 8;10(4):e033516. [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
20. Malterud K. Qualitative research: standards, challenges, and guidelines. Lancet 2001 Aug 11;358(9280):483-488. [doi: 10.1016/S0140-6736(01)05627-6] [Medline: 11513933]
21. Norekvål TM, Allore HG, Bendz B, et al. Rethinking rehabilitation after percutaneous coronary intervention: a protocol of a multicentre cohort study on continuity of care, health literacy, adherence and costs at all care levels (the CONCARDPCI). BMJ Open 2020 Feb 12;10(2):e031995. [doi: 10.1136/bmjopen-2019-031995] [Medline: 32054625]
22. Brørs G, Pettersen TR, Hansen TB, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. BMC Health Serv Res 2019 Jun 10;19(1):364. [doi: 10.1186/s12913-019-4106-1] [Medline: 31182100]
23. Valaker I, Norekvål TM, Råholm MB, et al. Continuity of care after percutaneous coronary intervention: the patient's perspective across secondary and primary care settings. Eur J Cardiovasc Nurs 2017 Jun;16(5):444-452. [doi: 10.1177/1474515117690298] [Medline: 28111970]
24. Hjertvikrem N, Brørs G, Instenes I, et al. Use of health services and perceived need for information and follow-up after percutaneous coronary intervention. BMC Res Notes 2024 Jan 5;17(1):20. [doi: 10.1186/s13104-023-06662-y] [Medline: 38183067]
25. Dominello A, Cheek C, Clay-Williams R. The use of personas to involve consumers in healthcare co-design: a scoping review. BMC Health Serv Res 2025 May 6;25(1):649. [doi: 10.1186/s12913-025-12786-7] [Medline: 40329348]
26. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YL, Kenny LC, Adam F. An integrated patient journey mapping tool for embedding quality in healthcare service reform. J Decis Syst 2016 Jun 10;25(sup1):354-368. [doi: 10.1080/12460125.2016.1187394]
27. Miaskiewicz T, Kozar KA. Personas and user-centered design: how can personas benefit product design processes? Des Stud 2011 Sep;32(5):417-430. [doi: 10.1016/j.destud.2011.03.003]
28. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
29. Stossel LM, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J Gen Intern Med 2012 Sep;27(9):1165-1170. [doi: 10.1007/s11606-012-2046-0] [Medline: 22528620]
30. Noushad B, Van Gerven PWM, de Bruin ABH. Twelve tips for applying the think-aloud method to capture cognitive processes. Med Teach 2024 Jul;46(7):892-897. [doi: 10.1080/0142159X.2023.2289847] [Medline: 38071621]
31. Brooke J. SUS—a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry, 1st edition: CRC Press; 1996:189-194.
32. Reshma SG, Mohan Kumar HP, Manu AG. Smoke test execution in software application testing. 2022 4th Int Conf Emerg Res Electron Comput Sci Technol 2022:1-7. [doi: 10.1109/ICERECT56837.2022.10059686]

XSL•FO
RenderX

33. Stevenson R, Burnell D, Fisher G. The minimum viable product (MVP): theory and practice. J Manage 2024 Nov;50(8):3202-3231. [doi: 10.1177/01492063241227154]
34. Wiarda M, Giannelos K, Schuerz S, Reber B, Doorn N. Ethics framework and guidelines for participatory processes in the activities of research funding organizations. : Zenodo; 2024 URL: https://zenodo.org/records/11505444 [accessed 2026-08-20]
35. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). : European Union; 2016 URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2026-02-01]
36. Cranen K, Groothuis-Oudshoorn CG, Vollenbroek-Hutten MM, IJzerman MJ. Toward patient-centered telerehabilitation design: understanding chronic pain patients' preferences for web-based exercise telerehabilitation using a discrete choice experiment. J Med Internet Res 2017 Jan 20;19(1):e26. [doi: 10.2196/jmir.5951] [Medline: 28108429]
37. Kizilkilic SE, Ramakers W, Falter M, et al. A digitally-supported shared decision making approach for patients during cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2025 Aug 28:zwaf537. [doi: 10.1093/eurjpc/zwaf537] [Medline: 40874533]
38. Gibson I, Jennings C, Neubeck L, et al. Using a digital health intervention "INTERCEPT" to improve secondary prevention in coronary heart disease (CHD) patients: protocol for a mixed methods non-randomised feasibility study. HRB Open Res 2023;6:43. [doi: 10.12688/hrbopenres.13781.2] [Medline: 38414839]
39. Golbus JR, Lopez-Jimenez F, Barac A, et al. Digital technologies in cardiac rehabilitation: a science advisory from the American Heart Association. Circulation 2023 Jul 4;148(1):95-107. [doi: 10.1161/CIR.0000000000001150] [Medline: 37272365]
40. Lundell S, Toots A, Sönnerfors P, Halvarsson A, Wadell K. Participatory methods in a digital setting: experiences from the co-creation of an eHealth tool for people with chronic obstructive pulmonary disease. BMC Med Inform Decis Mak 2022 Mar 18;22(1):68. [doi: 10.1186/s12911-022-01806-9] [Medline: 35303895]
41. Kraselnik A, McGowan H, Amiali IA, et al. Factors influencing the uptake of digital health interventions for cardiovascular disease prevention among healthcare providers: a systematic review. Glob Heart 2026;21(1):45. [doi: 10.5334/gh.1561] [Medline: 42253600]
42. Leask CF, Sandlund M, Skelton DA, et al. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. Res Involv Engagem 2019;5(1):2. [doi: 10.1186/s40900-018-0136-9] [Medline: 30652027]
43. Pepera G, Antoniou V, Su JJ, Lin R, Batalik L. Comprehensive and personalized approach is a critical area for developing remote cardiac rehabilitation programs. World J Clin Cases 2024 Apr 26;12(12):2009-2015. [doi: 10.12998/wjcc.v12.i12.2009] [Medline: 38680265]
44. Humayoun SR, Dubinsky Y, Catarci T. A three-fold integration framework to incorporate user-centered design into agile software development. In: Kurosu M, editor. Human Centered Design Second International Conference, HCD 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings: Springer:55-64. [doi: 10.1007/978-3-642-21753-1_7]
45. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74. [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
46. Jansen BJ, Salminen JO, Jung SG. Data-driven personas for enhanced user understanding: combining empathy with rationality for better insights to analytics. Data Inf Manag 2020 Mar;4(1):1-17. [doi: 10.2478/dim-2020-0005]
47. Yang Z, Xu L, Gao Y, Zhang C, Wang A. Tailored personas for self-management in home-based cardiac rehabilitation for patients with coronary heart disease: a qualitative study. Int J Nurs Stud 2025 Mar;163:105000. [doi: 10.1016/j.ijnurstu.2025.105000] [Medline: 39854909]
48. Lee EW, McCloud RF, Viswanath K. Designing effective eHealth interventions for underserved groups: five lessons from a decade of eHealth intervention design and deployment. J Med Internet Res 2022 Jan 7;24(1):e25419. [doi: 10.2196/25419] [Medline: 34994700]
49. Adewusi BA, Adekunle BI, Mustapha SD, Uzoka AC. Advances in AI-augmented user experience design for personalized public and enterprise digital services. Int J Adv Multidisc Res Stud 2023;3(6):1910-1929. [doi: 10.62225/2583049X.2023.3.6.4408]

Abbreviations

ALM: application life cycle management
CAD: coronary artery disease
CONCARD^{PCI}: continuity of care after percutaneous coronary intervention
CR: cardiac rehabilitation
CVD: cardiovascular disease
GP : general practitioner

XSL•FO
RenderX

GUIDED: Guidance for Reporting Intervention Development Studies in Health Research
ICT: information and communication technology
MRC: Medical Research Council
MVP: minimal viable product
NORMEN: Norwegian Code of Conduct for Information Security and Data Protection in the Health Care Sector
PaaS: platform as a service
PCI: percutaneous coronary intervention
PI: principal investigator
PROCARD: Patient-Reported Outcomes in Cardiology
SUS: System Usability Scale

Edited by I Steenstra; submitted 13.Feb.2026; peer-reviewed by K Kitagaki, ST Kulnik; revised version received 05.Aug.2026; accepted 05.Aug.2026; published 01.Sep.2026.

Please cite as:
Nordfonn OK, Johnsen Ramstad K, Roed Pettersen T, Hjertvikrem N, Aamot Aksetøy IL, Agcaoili Borge AJ, Gerdts E, Hesthammer JG, Hole T, Instenes I, Isaksen A, Krumsvik OA, Landaas M, Moltu C, Rotevatn S, Schjøtt J, Sommersten CH, Tryti JO, Norekvål TM Co-Creation and Usability Testing of a Digital Cardiac Rehabilitation Program (eCardiacRehab): Design and Development Study JMIR Cardio 2026;10:e93398
URL: https://cardio.jmir.org/2026/1/e93398
doi:10.2196/93398

© Oda Karin Nordfonn, Kristin Johnsen Ramstad, Trond Roed Pettersen, Nina Hjertvikrem, Inger-Lise Aamot Aksetøy, Alf Johannes Agcaoili Borge, Eva Gerdts, Jan Gunnar Hesthammer, Torstein Hole, Irene Instenes, Anita Isaksen, Ole Andreas Krumsvik, Merethe Landaas, Christian Moltu, Svein Rotevatn, Jan Schjøtt, Cathrine Horn Sommersten, Jan Ove Tryti, Tone Merete Norekvål. Originally published in JMIR Cardio (https://cardio.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of Wrist-Worn Activity Tracker Accuracy for Measuring Sedentary Behavior in Outpatients With Cardiovascular Disease: Validation Study

Koichi Naito[1,2], PT, MSc, PhD; Kazuhiro P Izawa[3], PT, MSc, PhD; Noriaki Maeda[4], PT, MSc, PhD; Yuya Kasai[2], PT; Haruna Tani[5], BSc; Misaki Kondo[1], BSc; Hajime Iwama[2], MD, PhD

[1]Department of Physical Therapy, Faculty of Medical Science, Nagoya Aoi University, 3-40 Shioji-cho, Mizuho-ku, Nagoya, Japan

[2]Iwama Cardiovascular and Dental Clinic for Prevention and Care, Kitakatsuragi-gun, Japan

[3]Department of Health Sciences, Graduate School of Medicine, Kobe University, Kobe, Japan

[4]Faculty of Sport Sciences, Waseda University, Saitama, Japan

[5]SapplyM, Inc., Tokyo, Japan

Corresponding Author:
Koichi Naito, PT, MSc, PhD
Department of Physical Therapy, Faculty of Medical Science, Nagoya Aoi University, 3-40 Shioji-cho, Mizuho-ku, Nagoya, Japan

Abstract

Background: Wearable monitoring of sedentary behavior is increasingly used in digital cardiovascular care, but the validity of consumer-grade wrist-worn trackers in older patients with cardiac issues is unclear.

Objective: This study aimed to evaluate the agreement between sedentary time measured by a consumer-grade wrist-worn tracker (Fitbit Inspire 3; Fitbit LLC) and a research-grade waist-worn accelerometer (ActiveStyle Pro HJA-750C; Omron Healthcare Co, Ltd) used as the comparator, in Japanese outpatients with cardiac issues under free-living conditions.

Methods: Fifty-one outpatients with heart failure (28/51, 55%) or ischemic heart disease (23/51, 45%; mean age 78.4, SD 9.0 years; 25/51, 49% male) simultaneously wore both devices for 7 consecutive days. Sedentary time was assessed within a standardized 12-hour daytime window (9 AM to 9 PM) at 1-minute epochs. Agreement was examined using Pearson correlation, Bland-Altman analysis (after confirming normality of the paired differences), Lin concordance correlation coefficient (CCC), and Cohen kappa for distribution-based tertile classification. A constant bias correction was applied, with leave-one-out cross-validation (LOOCV) used as an internal-consistency check on the stability of the sample mean bias. Moderators of the bias were explored using multivariable regression.

Results: Mean sedentary time within the 12-hour window was 497.9 (SD 86.5) minutes for the comparator and 559.9 (SD 84.3) minutes for the Fitbit. The devices were strongly correlated (r=0.905, 95% CI 0.838 - 0.945; P<.001). The Fitbit showed a significant fixed bias of 62.0 minutes (95% CI 51.5 - 72.5; P<.001) with no proportional bias (slope −0.03; P=.67); the 95% limits of agreement were wide (−11.3 to 135.2 minutes). Uncorrected concordance was moderate (CCC=0.71, 95% CI 0.57 - 0.80) and improved to high after subtracting the 62-minute bias (CCC=0.904, 95% CI 0.84 - 0.94); LOOCV indicated that this in-sample improvement was internally stable (out-of-sample CCC=0.90). No candidate factor (sex, age, diagnosis, and musculoskeletal disorder) significantly moderated the bias (model P=.12). Tertile classification showed moderate agreement (κ=0.588, 95% CI 0.405 - 0.772; 37/51, 72.5% agreement), with all misclassifications between adjacent categories.

Conclusions: In older Japanese outpatients with cardiac issues, a consumer-grade wrist-worn tracker overestimated sedentary time by approximately 62 minutes within a standardized 12-hour daytime window compared with a waist-worn comparator, with fixed but no proportional bias. A simple constant correction markedly improved group-level concordance (cross-validated CCC≈0.90), supporting calibrated use of consumer wearables for scalable monitoring in ambulatory cardiac care. However, wide limits of agreement constrain individual-level precision, and the population-specific correction factor requires independent validation before transfer to other settings.

(JMIR Cardio 2026;10:e91104) doi:10.2196/91104

KEYWORDS

sedentary behavior; wearable devices; consumer wearable; fitbit; validation study; bland-altman; concordance correlation coefficient; cardiac rehabilitation

XSL•FO
RenderX

Introduction

Sedentary behavior (SB)—any waking behavior performed in a sitting, reclining, or lying posture with an energy expenditure of ≤1.5 metabolic equivalents (METs)—has become recognized as a powerful, modifiable risk factor for cardiovascular morbidity and mortality [1,2]. A 2024 systematic review and dose-response meta-analysis of 19 prospective cohorts comprising 1,473,354 individuals and 60,526 cardiovascular events found that participants in the highest SB category had a 30% higher risk of fatal and nonfatal cardiovascular disease (CVD) compared with the lowest category (relative risk [RR] 1.29, 95% CI 1.22 - 1.37). Each additional hour of sedentary time was associated with a 5% increase in CVD risk (RR 1.05, 95% CI 1.02 - 1.07). The dose-response curve exhibited a J-shaped pattern, and replacing one hour of sitting with light-intensity physical activity (LPA) lowered CVD risk by 20% (RR 0.84, 95% CI 0.73-0.97) [3]. These quantitative estimates underline the clinical importance of reducing and breaking up prolonged sitting [4,5]. Global public health guidance also explicitly highlights reducing sedentary behavior alongside increasing physical activity across the life course [6].

A recent Japanese study of 92 outpatients undergoing cardiac rehabilitation measured with a triaxial waist-worn accelerometer reported average SB of 663 (SD 118) minutes/day (≈11 hours) and moderate-to-vigorous physical activity (MVPA) of just 25 minutes/day; men sat 64 minutes/day more than women and performed 22% less light-intensity activity [7]. Such findings underscore the importance of sex-specific strategies to break up SB in Asian settings where population aging is pronounced [8,9].

Emerging research emphasizes that wearable device–based measurement and intervention targeting sedentary behavior are critical components of secondary prevention strategies in patients with coronary artery disease (CAD). Continuous monitoring of sedentary time allows for the identification of at-risk individuals, while targeted interventions informed by wearable analytics can support behavioral modification and contribute to improved cardiovascular outcomes [10-12]. Research-grade triaxial accelerometers, such as the waist-worn ActiveStyle Pro (Omron Healthcare Co., Ltd.), have demonstrated high validity for estimating METs compared with the Douglas bag method and are considered effective for sedentary behavior measurement [13]. Nevertheless, maintaining compliance with device wear over prolonged monitoring periods remains a challenge.

Inexpensive wrist-worn consumer trackers—most prominently the Fitbit (Fitbit LLC) series—are ubiquitous, integrate seamlessly with smartphone dashboards, and are attractive for remote monitoring [14,15]. Yet their proprietary algorithms were engineered primarily for step detection; recent investigations in healthy adults show systematic overestimation of SB stemming from arm movements [16-18], and wrist placement cannot directly capture postural transitions that define sedentary behavior, requiring inference from movement patterns that can be prone to misclassification [19]. Their validity in clinical populations therefore remains uncertain [20].

Thus, we aimed to evaluate the agreement between sedentary time estimates derived from a consumer-grade wrist-worn device (Fitbit) and a research-grade waist-worn accelerometer (ActiveStyle Pro) in Japanese outpatients with cardiac issues under free-living conditions. We used Bland-Altman plots to characterize systematic and proportional bias [21], calculated Lin concordance-correlation coefficient (CCC) [22], and assessed tertile classification agreement using Cohen kappa [23]. Prior validation studies [24-28] comparing wrist-worn Fitbits with research-grade monitors report mixed bias for sedentary time, with differences ranging from modest underestimation to overestimation depending on the comparator (ActiGraph; ActiGraph LLC vs activPAL; PAL Technologies Ltd), wear location, population, and data processing. Accordingly, we hypothesized the presence of systematic bias without prespecifying its direction and used Bland-Altman analysis to characterize both fixed and proportional components, also testing whether a constant offset improved concordance.

By clarifying the magnitude and nature of measurement error in this clinically important population—and situating our results alongside emerging Japanese data highlighting high SB levels and sex differences [7]—our findings aim to guide researchers selecting wearable sensors for SB trials and to assist clinicians interpreting tracker-derived metrics when delivering behavior-change counseling in ambulatory cardiac care.

Methods

Study Design

This cross-sectional validation study was conducted at a single cardiology outpatient clinic in Nara Prefecture, Japan, between October 2022 and March 2023.

Patient Characteristics

Fifty-one consecutive outpatients with CAD or heart failure who were enrolled in the outpatient cardiac rehabilitation program were recruited for this study. Inclusion criteria were as follows: (1) diagnosed with ischemic heart disease or heart failure, (2) aged ≥18 years, (3) able to walk independently without assistive devices, (4) willing to wear both activity monitors simultaneously for 7 consecutive days, and (5) capable of providing informed consent. Exclusion criteria included (1) acute decompensated heart failure requiring hospitalization, (2) unstable angina or recent acute coronary syndrome within 4 weeks, (3) severe cognitive impairment precluding informed consent, (4) skin conditions preventing device attachment, and (5) planned hospitalization or major procedure during the monitoring period.

Demographic and clinical data were extracted from medical records, including age, sex, height, weight, BMI, primary cardiac diagnosis (heart failure vs ischemic heart disease), comorbidities (hypertension, dyslipidemia, diabetes mellitus, respiratory disease, renal disease, cerebrovascular disease, musculoskeletal disorders, malignancy), and current medications (beta-blockers, hyperpolarization-activated cyclic nucleotide [HCN]-gated channel blockers, angiotensin-converting enzyme [ACE] inhibitors or angiotensin receptor blockers [ARBs], mineralocorticoid receptor antagonists, angiotensin

XSL•FO
RenderX

receptor–neprilysin inhibitor [ARNI], sodium-glucose cotransporter-2 [SGLT2] inhibitors, calcium channel blockers, nitrates, diuretics, statins). New York Heart Association (NYHA) functional class was not collected because the cohort included both patients with heart failure and patients with ischemic heart disease, for whom NYHA classification was not uniformly applicable. Musculoskeletal disorders were extracted from the medical records as an overall comorbidity category. No participant had a documented upper-limb musculoskeletal disorder that was expected to limit arm movement or affect wrist-worn device measurements.

Sedentary Time

Participants simultaneously wore 2 activity monitors for 7 consecutive days: a research-grade triaxial accelerometer (ActiveStyle Pro HJA-750C) positioned on the waist and a consumer-grade wrist-worn tracker (Fitbit Inspire 3) on the nondominant wrist. The ActiveStyle Pro has been validated for measuring METs against the Douglas bag method [13] and its comparability with other commonly used activity monitors has been examined in free-living contexts [29,30]. The ActiveStyle Pro is widely used in sedentary behavior research [7,31]. The Fitbit Inspire 3 represents the latest generation of consumer wearables with integrated smartphone connectivity.

Participants were instructed to wear both devices continuously during waking hours (from awakening until bedtime) but were permitted to remove them during water-based activities (bathing and swimming). Detailed written and verbal instructions were provided regarding proper device placement and daily charging procedures. Participants maintained a simple daily log recording wake and sleep times and any periods when devices were removed.

Sedentary time was defined as any waking behavior performed in a sitting, reclining, or lying posture with an energy expenditure ≤1.5 METs, consistent with international consensus definitions [1]. For analysis, we extracted sedentary time data from both devices for the period between 9 AM and 9 PM to standardize the measurement window across participants and minimize the influence of sleep-related recordings. Data were recorded at 1-minute epochs. Daily sedentary time was calculated as the sum of all minutes classified as sedentary behavior within the standardized 12-hour measurement window.

To be included in the final analysis, participants were required to have at least 4 valid days of data, defined as days with ≥10 hours of concurrent wear time for both devices within the 9 AM to 9 PM measurement window. Mean sedentary time within the standardized 12-hour window across all valid days was calculated for each participant for both the ActiveStyle Pro and Fitbit devices.

Fitbit Data Extraction and Sedentary Classification

Minute-level activity data were retrieved from the Fitbit Web API (Activity Intraday time series), which returns data at 1-minute intervals. As the intensity index for each epoch, we used the MET value returned with the calories resource. Because the Fitbit API expresses this value as METs multiplied by 10 (eg, a returned value of 15 corresponds to 1.5 METs), we defined an epoch as sedentary when the returned MET value was below 15 (ie, <1.5 METs) and the step count for the same minute was zero. This operationalization aligns the Fitbit-derived classification with the international ≤1.5 MET definition of sedentary behavior applied to the ActiveStyle Pro comparator. We emphasize, however, that this classification is an estimate based on the activity intensity and step information provided by the device and does not constitute a direct measurement of posture; this difference in how “sedentary” is operationalized between the 2 devices is itself a likely contributor to the observed bias.

Nonwear Time

Minutes for which the Fitbit returned no detected heart rate were treated as nonwear and excluded from analysis. For the ActiveStyle Pro, a period of 60 or more consecutive minutes of zero counts was treated as nonwear. Valid days required at least 10 hours of concurrent wear time for both devices within the 9 AM to 9 PM window, and participants with at least 4 valid days were retained, as described above.

Daytime Sleep and Naps

Participants were instructed to record any daytime napping in their paper daily log. No participant reported a daytime nap during the monitoring period; therefore, daytime sleep did not contribute to the sedentary tallies of either device, and any device-specific differences in nap handling could not have confounded the present comparison.

Fitbit Companion App Configuration

The wrist-placement setting in the Fitbit companion app was configured to match the actual placement of the device on the participant’s nondominant wrist, ensuring that the device’s motion-processing sensitivity corresponded to the physical wear location and removing altered handedness settings as a potential source of systematic error.

Statistical Analysis

Continuous variables are presented as mean (SD) or median (IQR) according to the normality of their distribution, assessed with the Shapiro-Wilk test. Categorical variables are presented as frequencies and percentages. Because n <100, percentages are reported without decimal places, and absolute numbers are reported alongside percentages.

Agreement between the ActiveStyle Pro comparator and the Fitbit was evaluated with several complementary analyses. Pearson correlation quantified the strength of the linear relationship. Bland-Altman analysis was used to assess systematic bias and to calculate the 95% limits of agreement [21]. Because the construction of parametric limits of agreement assumes that the paired differences are normally distributed, we first verified this assumption with the Shapiro-Wilk test applied to the differences. Fixed bias was tested with a one-sample *t* test of the paired differences, and proportional bias was examined by regressing the differences on the means of the two measurements.

Lin CCC was calculated to evaluate overall agreement, combining precision and accuracy, with 95% CIs obtained by bootstrap resampling [22]. The CCC ranges from −1 to +1, with values closer to 1 indicating better agreement. Categorical

XSL•FO
RenderX

agreement was assessed by classifying each participant into distribution-based tertiles (lower, middle, and upper third) of sedentary time separately for each device and computing the Cohen kappa coefficient for the resulting categories [23].

Where Bland-Altman analysis indicated significant fixed bias, we conducted a secondary analysis subtracting the mean bias from the Fitbit values and recomputing agreement. As an internal consistency check on the stability of the sample mean bias, we recomputed the correction using leave-one-out cross-validation (LOOCV): for each participant, the mean bias was estimated from the remaining participants and applied to that held-out participant, and the concordance of the cross-validated, bias-corrected Fitbit values with the comparator was then assessed. As a sensitivity analysis, we repeated this resampling procedure using repeated random 50/50 splits (Monte Carlo cross-validation, 2000 repetitions).

As an exploratory, hypothesis-generating analysis of whether the bias was moderated by participant characteristics, we fitted a single multivariable linear regression of the individual difference (Fitbit−ActiveStyle Pro) on sex, age, primary cardiac diagnosis (heart failure vs ischemic heart disease), and presence of a musculoskeletal disorder. Because the study was not powered for subgroup detection, a single adjusted model was used rather than multiple stratified tests.

All analyses were performed in R version 4.1.3 (R Foundation for Statistical Computing); statistical significance was set at $P<.05$.

Ethical Considerations

This study was approved by the Ethics Review Committee of Nagoya Aoi University (formerly Nagoya Women's University; approval number 2022‐33) and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment. To protect privacy and confidentiality, data collection and data analysis were carried out by different personnel, and analyses were performed on deidentified data. No compensation was provided to participants. No images or other materials permitting identification of individual participants are included in this manuscript or its supplementary files.

Results

Patient Characteristics

A total of 51 outpatients (25/51, 49% male; mean age 78.4, SD 9.0 years) were enrolled in the study. Participant characteristics are summarized in Table 1. The mean BMI was 23.8 (SD 4.6) kg/m². The primary cardiac diagnosis was heart failure in 55% (28/51) of participants and ischemic heart disease in 45% (23/51). The cohort had a high burden of cardiovascular risk factors and comorbidities: hypertension (44/51, 86%), dyslipidemia (36/51, 71%), diabetes mellitus (12/51, 24%), renal disease (20/51, 39%), and musculoskeletal disorders (18/51, 35%). Most participants were receiving evidence-based cardiovascular medications, including statins (34/51, 67%), ARNI (31/51, 61%), beta-blockers (24/51, 47%), calcium channel blockers (21/51, 41%), and SGLT2 inhibitors (31%, 16/51; Table 1).

XSL•FO
RenderX

Table . Baseline characteristics of study participants (N=51).

Characteristics	Value
Demographics	
Age (years), mean (SD)	78.4 (9.0)
Sex (male), n (%)	25 (49)
Height (cm), mean (SD)	158.0 (8.6)
Weight (kg), mean (SD)	59.3 (12.8)
BMI (kg/m²), mean (SD)	23.8 (4.6)
Primary cardiac diagnosis, n (%)	
Heart failure	28 (55)
Ischemic heart disease	23 (45)
Comorbidities, n (%)	
Hypertension	44 (86)
Dyslipidemia	36 (71)
Diabetes mellitus	12 (24)
Renal disease	20 (39)
Musculoskeletal disorders	18 (35)
Respiratory disease	6 (12)
Cerebrovascular disease	5 (10)
Malignancy	4 (8)
Medications, n (%)	
Statins	34 (67)
ARNI[a]	31 (61)
Beta-blockers	24 (47)
Calcium channel blockers	21 (41)
SGLT2[b] inhibitors	16 (31)
ARB[c]	11 (22)
HCN[d] channel blockers	11 (22)
Mineralocorticoid receptor antagonists	7 (14)
ACE[e] inhibitors	3 (6)
Nitrates	2 (4)
Diuretics	2 (4)

[a]ARNI: angiotensin receptor–neprilysin inhibitor.

[b]SGLT2: sodium-glucose cotransporter-2.

[c]ARB: angiotensin receptor blocker.

[d]HCN: hyperpolarization-activated cyclic nucleotide–gated.

[e]ACE: angiotensin-converting enzyme.

All participants successfully completed the 7-day monitoring period with both devices. The median number of valid measurement days was 7 (IQR 7 - 7), with all participants achieving at least 4 valid days of concurrent device wear. Mean daily wear time within the standardized measurement window (9 AM to 9 PM) was 11.8 (SD 0.4) hours, indicating high compliance with the monitoring protocol.

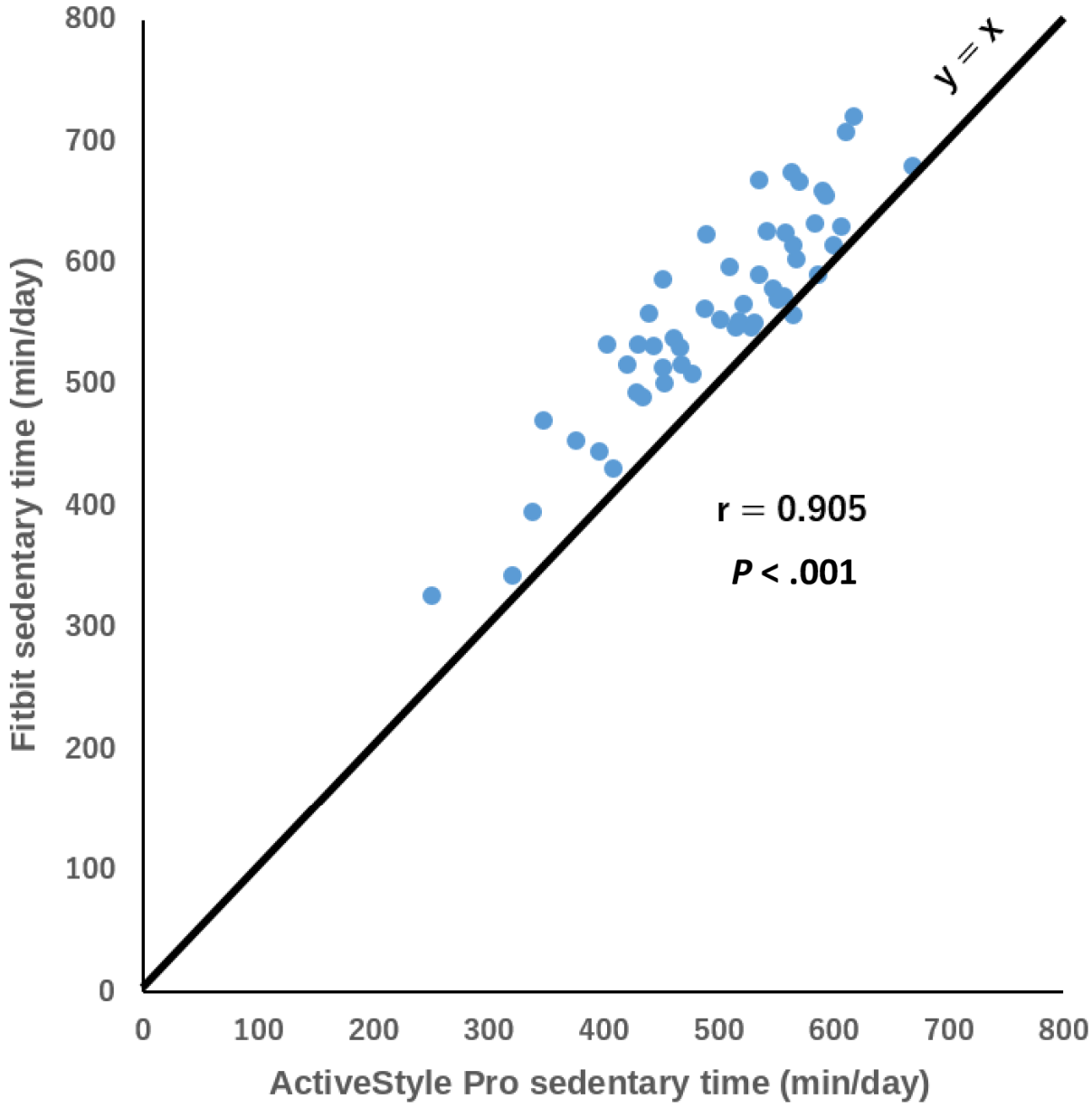
Agreement Between ActiveStyle Pro and Fitbit Measurements

Mean sedentary time within the standardized 12-hour window (9 AM to 9 PM) was 497.9 (SD 86.5) minutes for the ActiveStyle Pro comparator and 559.9 (SD 84.3) minutes for the Fitbit. A strong positive correlation was found between the two devices (Pearson r=0.905, 95% CI 0.838 - 0.945; P<.001; Figure 1).

XSL•FO
RenderX

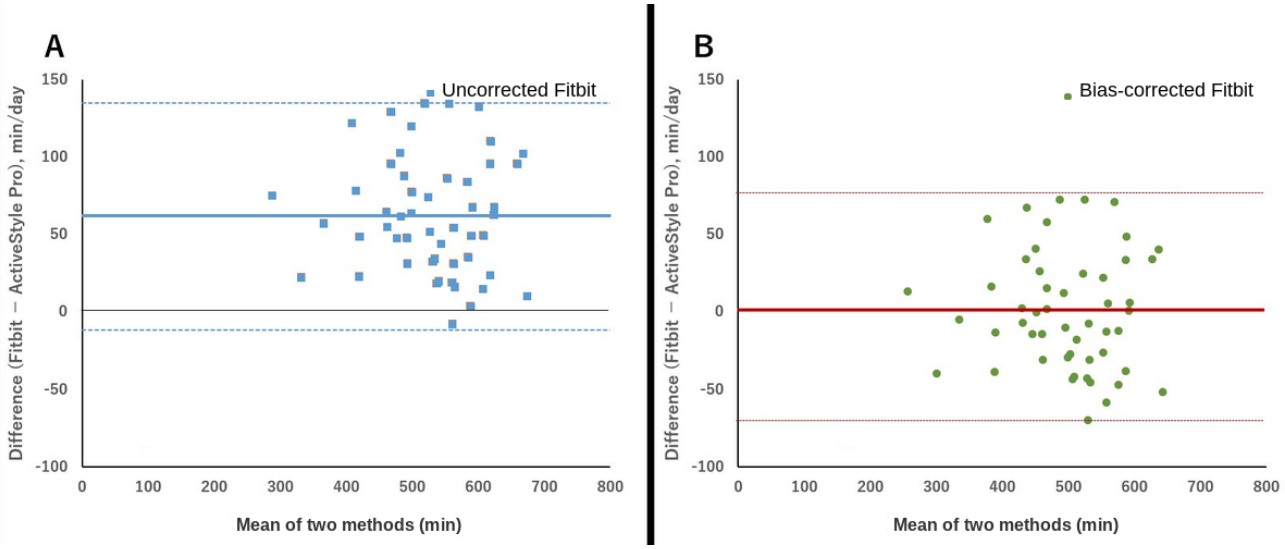
The paired differences were normally distributed (Shapiro-Wilk W=0.97; *P*=.16), supporting the use of parametric limits of agreement. Bland-Altman analysis revealed a significant fixed bias of 62.0 minutes (95% CI 51.5 - 72.5; *P*<.001), indicating that the Fitbit systematically overestimated sedentary time relative to the comparator. The 95% limits of agreement were wide, ranging from −11.3 to 135.2 minutes, indicating substantial interindividual variability in the magnitude of disagreement. Regression of the differences on the means showed a significant intercept of 76.4 minutes (95% CI 7.7 - 145.2; *P*=.03) and a nonsignificant slope of −0.03 (95% CI −0.16 to 0.10; *P*=.67), confirming fixed bias without proportional bias; that is, the degree of overestimation was approximately constant across the range of sedentary time (Figure 2).

Figure 1. Scatter plot of mean sedentary time within the standardized 12-hour window (9 AM to 9 PM) measured by Fitbit Inspire 3 vs ActiveStyle Pro (HJA-750C) in outpatients with cardiac issues (N=51). Each point represents one participant's mean sedentary time across valid days. Pearson correlation coefficient is shown.



XSL•FO
RenderX

Figure 2. Bland-Altman plots comparing sedentary time between Fitbit Inspire 3 and ActiveStyle Pro (HJA-750C) in outpatients with cardiac issues (N=51). The y-axis shows the difference (Fitbit−ActiveStyle Pro) and the x-axis shows the mean of the two methods. The horizontal line at y=0 indicates zero difference (perfect agreement). Solid lines indicate the mean difference (bias), and dashed lines indicate the 95% limits of agreement. A: uncorrected Fitbit estimates. B: bias-corrected Fitbit estimates obtained by subtracting 62 minutes per 12-hour window (9 AM to 9 PM) from Fitbit values.



When sedentary time was classified into distribution-based tertiles computed separately for each device (ActiveStyle Pro boundaries: lower third <458 minutes, middle third 458 - 549 minutes, upper third >549 minutes; Fitbit boundaries: lower third <533 minutes, middle third533 - 593 minutes, upper third >593 minutes), the Cohen kappa coefficient was 0.588 (95% CI 0.405 - 0.772), indicating moderate categorical agreement. The overall agreement rate was 72.5% (37/51 participants), with an agreement of 82.4% (14/17), 58.8% (10/17), and 76.5% (13/17) in the lower, middle, and upper tertiles, respectively. All 14 misclassifications (27.5%) occurred between adjacent tertiles, with no participant misclassified by two or more categories. The middle tertile showed the lowest agreement, indicating that classification was most difficult near category boundaries. The mean absolute difference in measured sedentary time among misclassified cases was 70.3 (SD 49.7) minutes.

After subtracting 62 minutes from the Fitbit values to correct for the fixed bias, regression of the corrected differences on the means showed no significant fixed bias (intercept=13.6 minutes, 95% CI −51.3 to 78.4; *P*=.68) while proportional bias remained absent (slope=−0.03; *P*=.67). The in-sample CCC improved from 0.71 (95% CI 0.57 - 0.80) to 0.904 (95% CI 0.84 - 0.94). To assess the internal stability of the in-sample correction, we used LOOCV: the out-of-sample CCC was 0.901, essentially identical to the in-sample value (shrinkage=0.004), and the per-fold bias estimate was highly stable (range 60.5 - 63.4 minutes; SD 0.7). Repeated random 50/50 splits (Monte Carlo cross-validation) yielded a mean out-of-sample concordance of 0.89 (95% interval 0.81 - 0.94). This reflects the stability of the sample mean within the present dataset and does not by itself establish that the correction factor generalizes to other samples (see Study limitations). Notwithstanding this group-level improvement, the limits of agreement for the cross-validated corrected values remained wide (approximately −74.7 to 74.7 minutes), indicating that correction reduces systematic bias at the group level but does not resolve individual-level imprecision (Figure 2B).

In the exploratory multivariable model of the individual difference (Fitbit−ActiveStyle Pro), none of the candidate characteristics significantly moderated the bias. The adjusted estimates were −13.9 minutes for male versus female sex (95% CI −34.4 to 6.5; *P*=.18), +0.7 minutes per additional year of age (95% CI −0.5 to 1.9; *P*=.25), +17.4 minutes for heart failure versus ischemic heart disease (95% CI −3.7 to 38.5; *P*=.10), and −6.7 minutes for the presence of a musculoskeletal disorder (95% CI −28.0 to 14.5; *P*=.53). The overall model explained little of the variance and was not statistically significant (R^2=0.15; model *P*=.12). Although heart failure versus ischemic heart disease and older versus younger age appeared associated with the bias when examined in isolation, neither remained significant after mutual adjustment, consistent with confounding between age and diagnosis and with multiple comparisons. However, this model was underpowered (N=51; model *P*=.12; R^2=0.15), and these null results should not be read as evidence that clinically relevant subgroup differences are absent. In particular, the adjusted estimate for heart failure versus ischemic heart disease (+17.4 minutes, 95% CI −3.7 to 38.5 minutes) remains compatible with a difference large enough to matter relative to the overall 62-minute bias. We therefore regard these analyses as hypothesis-generating only and do not conclude that a single population-level correction is definitively adequate across clinically important subgroups (see Study limitations).

Discussion

Summary of Main Findings

This study evaluated the validity of a consumer-grade wrist-worn activity tracker (Fitbit Inspire 3) for measuring sedentary time in Japanese outpatients with cardiac issues, using a research-grade waist-worn accelerometer (ActiveStyle Pro) as the comparator. We found that the Fitbit systematically overestimated sedentary time by approximately 62 minutes within a standardized 12-hour daytime window (9 AM to 9 PM; values are therefore not full-day estimates), with clear evidence of fixed bias but no proportional bias across the range of

XSL•FO
RenderX

sedentary behavior. Although uncorrected agreement was only moderate, applying a simple constant correction substantially improved concordance between devices. Taken together, these findings indicate that consumer-grade wrist-worn devices can provide clinically meaningful information on sedentary behavior in cardiac outpatients with cardiac issues when their systematic measurement error is explicitly acknowledged and appropriately calibrated.

Comparison With Previous Validation Studies

Our finding that the Fitbit overestimated sedentary time by approximately 62 minutes within the standardized 12-hour daytime window (9 AM to 9 PM) contrasts with some previous validation studies [24,25] that reported underestimation of sedentary time by wrist-worn devices. This discrepancy warrants careful consideration and may be explained by several factors related to our specific study population and context, while also revealing important insights into the mechanisms underlying device measurement error.

Population-Specific Movement Patterns

Population characteristics likely play a crucial role in explaining the direction of bias. Our study enrolled older outpatients with cardiac issues (mean age 78.4, SD 9.0 years) with heart failure or ischemic heart disease, whereas many previous validation studies examined younger, healthier adults. The mean sedentary time observed with the ActiveStyle Pro comparator was 498 minutes within the standardized 12-hour window (9 AM to 9 PM; approximately 8.3 hours). Extrapolated across typical waking hours, this is broadly consistent with the approximately 11 hours per day reported in recent Japanese data on rehabilitation patients with cardiac issues, although our measurement was confined to the standardized window. This underscores both the clinical relevance of accurate sedentary behavior measurement in this population and the distinctive activity profile that may influence device performance.

Older patients with cardiac issues may exhibit distinct movement patterns during daily activities. Specifically, even when standing or performing light activities such as cooking, housework, or slow walking, these patients may have minimal arm displacement due to age-related changes in gait, cardiovascular limitations restricting vigorous arm swing, or cautious movement patterns to avoid symptoms. Consequently, periods of standing or light activity with stationary arms may be misclassified by the wrist-worn device as sedentary time, leading to overestimation.

In contrast, previous studies [24,25] reporting underestimation may have examined populations where wrist movements during sedentary activities (eg, typing, using smartphones or tablets while sitting, and hand gestures during conversation) were sufficient to trigger activity classification, resulting in sedentary time being counted as active time. Furthermore, participants in these studies had mean ages of approximately 20 and 68 years, which are younger than those in the current study, suggesting that age-related differences in movement patterns may partly explain this discrepancy.

Taken together, these considerations suggest that the overestimation observed in older outpatients with cardiac issues reflects predictable interactions between age-related movement patterns and the limitations of wrist-based sensing, underscoring the need for population-specific validation when applying consumer wearables in clinical cardiac care.

Device Design and Algorithmic Factors

The consistent 62-minute overestimation reflects fundamental differences between wrist-worn and waist-worn device designs. Wrist placement cannot directly measure postural transitions (sitting-to-standing and lying-to-sitting) that are the defining feature of sedentary behavior. Instead, wrist devices must infer posture from arm movement patterns—an indirect approach prone to misclassification [19,25]. When individuals stand still or perform light activities with minimal arm displacement, the wrist device interprets the lack of movement as sedentary behavior, regardless of actual posture. Conversely, waist-worn devices like the ActiveStyle Pro capture whole-body movement and postural changes more directly through detection of trunk position and acceleration patterns.

Proprietary algorithms in consumer devices are optimized primarily for step counting and moderate-to-vigorous activity tracking, the metrics most valued by general consumers, rather than sedentary time classification. The Fitbit Inspire 3 represents a newer generation of consumer wearables with updated algorithms compared to devices examined in earlier studies. The threshold and decision rules for classifying sedentary versus active states may have been modified across device generations and may not be optimized for populations with limited mobility or distinct movement signatures.

Comparator Device and Methodological Considerations

Device and algorithm differences across validation studies must be considered. Our comparator was the ActiveStyle Pro, a waist-worn triaxial accelerometer validated against metabolic measurements (Douglas bag method), whereas previous studies often used the ActiGraph or activPAL (a thigh-worn inclinometer). Different comparator devices use distinct classification approaches, posture-based (activPAL, using inclinometry) versus movement-based (ActiGraph and ActiveStyle Pro, using acceleration thresholds), which may influence the magnitude and direction of observed bias. Direct comparison across validation studies is complicated by this methodological heterogeneity.

Additionally, activity context and measurement protocol may contribute to differences across studies. We measured sedentary time during a standardized 12-hour daytime window (9 AM to 9 PM) in free-living conditions. In older patients with cardiac issues, this window likely captures prolonged periods of sitting while watching television, reading, or resting—activities performed with relatively stationary arms. If healthy adults in previous studies engaged in more dynamic activities with frequent arm movements during nonsedentary time, wrist devices might better distinguish active from sedentary periods in those populations.

XSL•FO
RenderX

Implications of the Absence of Proportional Bias

Importantly, the absence of proportional bias in our data, confirmed by the nonsignificant regression slope (P=.67), indicates that the magnitude of overestimation remains constant regardless of whether individuals have low or high sedentary time. This finding has both mechanistic and practical implications.

Mechanistically, the consistent bias suggests that overestimation stems from systematic misclassification of specific activity types (low-movement standing and light activities) rather than from device calibration errors that would vary with activity volume. If the bias were due to gain or sensitivity errors, we would expect proportional bias where the error increases with higher sedentary time. Instead, the fixed bias pattern indicates that a relatively constant duration of nonsedentary activities (approximately 62 minutes within the 12-hour window; 9 AM to 9 PM) is being misclassified as sedentary across all participants.

Practically, this finding means that a simple constant correction can substantially improve agreement, as evidenced by the improvement in CCC from 0.71 to 0.904 after subtracting 62 minutes. This correction approach would not be appropriate if proportional bias were present.

Context Dependency of Device Validity

The mixed findings across validation studies, including our own, underscore a critical point: wearable device validity is population-specific and context-dependent. A device validated in one population cannot be assumed to perform equivalently in another, particularly when movement patterns and activity repertoires differ substantially. This context-dependency extends beyond population characteristics to include cultural factors (eg, floor-sitting in some Asian populations), occupational demands, and the specific activities that comprise daily routines.

Research Applications

For research applications, our results suggest that Fitbit devices should not be used interchangeably with research-grade accelerometers for primary outcome assessment without applying appropriate corrections or acknowledging measurement limitations. The moderate uncorrected CCC (0.71) indicates insufficient agreement for studies requiring precise absolute sedentary time measurements. However, the excellent agreement after constant correction (CCC=0.904) demonstrates that simple calibration may render consumer devices suitable for some research contexts.

Population-specific calibration should be performed before using consumer devices in research. Our 62-minute correction factor applies specifically to older Japanese outpatients with cardiac issues and should not be assumed to apply to other populations without validation. Researchers planning to use Fitbit or similar devices should consider conducting preliminary validation studies in their target population to determine appropriate correction factors.

The moderate categorical agreement (κ=0.588) when classifying participants into sedentary time tertiles suggests acceptable performance for epidemiological studies examining associations between categorical sedentary behavior levels and health outcomes. Importantly, all misclassifications occurred between adjacent categories with no extreme misclassifications, indicating that the device maintains relative ordering of individuals. This property may be sufficient for observational studies where precise quantification is less critical than rank-ordering or categorical stratification.

For intervention trials, where detecting within-person change is paramount, our cross-sectional data provide only indirect evidence of suitability. The absence of proportional bias suggests that changes in sedentary time should be measured with similar accuracy regardless of baseline levels, but longitudinal validation is needed to confirm this hypothesis and quantify measurement error for within-person change.

Consumer wearables may offer particular value in large-scale studies where research-grade devices are cost-prohibitive. With appropriate calibration and acknowledgment of limitations, devices like the Fitbit could enable sedentary behavior research in larger, more diverse samples, potentially improving external validity and generalizability of findings.

Study Limitations

Several limitations warrant consideration and point to important directions for future research.

Cross-Sectional Design and Temporal Stability

This was a cross-sectional validation study that assessed agreement between devices at a single time point. While we found strong between-person correlation, the study design does not directly evaluate the device's ability to track within-person changes over time. Test-retest reliability, day-to-day measurement error, and the minimal detectable change for individual monitoring remain unknown. Future longitudinal studies with repeated measurements over weeks to months are essential to establish whether the device can reliably detect behavioral changes within individuals—a critical requirement for clinical monitoring and intervention trials.

Responsiveness to intervention should be evaluated in prospective studies where sedentary time is intentionally modified through behavioral counseling or structured programs. Such studies would determine whether Fitbit-detected changes correspond to changes measured by research-grade devices and whether these changes predict clinical outcomes such as cardiovascular events, functional capacity, or quality of life.

Measurement Variability and Precision

The wide 95% limits of agreement (−11.3 to 135.2 minutes) indicate substantial interindividual variability in measurement error, corresponding to approximately ±73 minutes around the mean bias. The sources of this variability remain unclear but warrant investigation. Potential moderators include:

- Individual movement characteristics: arm movement patterns during activities, gait characteristics, and activity repertoire.
- Physiological factors: body composition (arm circumference, adiposity), cardiovascular functional capacity, and presence of tremor or dyskinesia.

XSL•FO
RenderX

- Device-related factors: wearing position on wrist, tightness of band, and skin temperature affecting sensor performance.
- Behavioral factors: types of activities performed, proportion of time in different postures, and occupation-related demands.

Future research should systematically evaluate these potential moderators to identify subgroups in whom device performance may be particularly strong or weak. Understanding sources of variability could inform personalized correction approaches or help clinicians identify patients for whom alternative monitoring methods may be preferable.

Our Fitbit classification required a step count of exactly zero for a minute to be counted as sedentary. Wrist-worn trackers are known to register "phantom" steps from nonambulatory arm movements such as gesturing, fidgeting, or hand activity performed while seated. Minutes containing such phantom steps would have been classified as nonsedentary even when the participant was in fact sitting, causing the device to under-count sedentary time in those minutes. The net 62-minute overestimation we observed may therefore mask bidirectional misclassification—overestimation when low-movement standing or light activity is misread as sedentary, partially offset by underestimation when seated, phantom-step minutes are misread as active—rather than a purely unidirectional error. The minute-level analyses needed to quantify these opposing components were beyond the scope of the present study and are a priority for future work.

A statistical caveat applies to our cross-validation of the bias correction. Subtracting a constant equal to the mean paired difference will, almost by construction, improve the concordance correlation coefficient, because it removes the location-shift (accuracy) component of disagreement while leaving the precision component unchanged. Furthermore, because the correction has a single parameter, a constant offset estimated as the sample mean, LOOCV re-estimates essentially the same mean at each fold, so the negligible shrinkage we observed reflects the stability of the sample mean within this dataset rather than evidence that the 62-minute correction factor is transportable to other samples or settings. Our cross-validation should therefore be interpreted as an internal-consistency check, not as external validation. Establishing the external validity of the correction factor requires deriving and testing it in independent populations.

Sedentary Patterns Beyond Total Time

Our analysis focused on total daily sedentary time within a standardized 12-hour window. We did not examine the validity of detecting sedentary bout duration or breaks in sedentary time—metrics that may be equally or more important for cardiovascular health than total volume. Emerging evidence suggests that prolonged uninterrupted sedentary bouts (eg, >30 or >60 minutes) and the frequency of breaks in sedentary time confer cardiovascular risks independent of total sedentary time. A recent meta-analysis found that replacing one hour of sitting with light-intensity physical activity lowered cardiovascular disease risk by 20%, highlighting the importance of breaking up prolonged sitting.

Future validation work should assess whether consumer devices can accurately identify:

- Bout duration: Can the device distinguish between 6 hours accumulated in 10-minute bouts versus 6 hours in uninterrupted sitting?
- Bout patterns: Can it detect breaks in sedentary time and classify their intensity (standing breaks vs light activity breaks)?
- Temporal patterns: Does the device accurately capture sedentary time distribution across the day (morning vs afternoon vs evening)?

These pattern-based metrics may be more clinically relevant than total volume for cardiovascular risk stratification and intervention targeting [32-36].

Generalizability Concerns

The study was conducted in a specific population of Japanese outpatients with cardiac issues with a mean age of approximately 78 (SD 9.0) years, and predominantly with heart failure or ischemic heart disease. Generalizability requires verification across:

- Age groups: younger patients with cardiac issues, middle-aged adults, and very older (>85 years).
- Cardiac conditions: valvular disease, arrhythmias, congenital heart disease, and posttransplant.
- Functional status: NYHA class I vs IV, different exercise capacities.
- Geographic and ethnic populations: Western populations, other Asian countries, and diverse ethnic groups.
- Noncardiac populations: healthy older adults, other chronic conditions (chronic obstructive pulmonary disease, diabetes, and chronic kidney disease).

We could not evaluate whether the measurement bias differed by NYHA functional class because NYHA class was not collected and was not uniformly applicable to all participants. Although musculoskeletal disorders were included as a broad comorbidity category, none of the participants had an upper-limb musculoskeletal disorder documented in the medical records that was expected to affect wrist-worn device measurements. Therefore, the influence of upper-limb impairment on wrist-worn device accuracy could not be evaluated in this cohort.

Device performance may differ substantially across these groups due to varying movement patterns, activity levels, and physiological constraints. The 62-minute correction factor derived from our study should not be applied to other populations without validation.

Although no participant had a documented upper-limb musculoskeletal disorder, 35% (18/51) of the cohort had a musculoskeletal disorder of some kind, and we did not record detailed gait characteristics. Because independent ambulation without an assistive device was an eligibility requirement, walking-aid users were not enrolled, and cane or walker use is therefore unlikely to account for misclassification in the present cohort. Within the enrolled cohort, lower-limb musculoskeletal disease and age-related gait alteration can nonetheless reduce natural arm swing during walking, even in individuals who

XSL•FO
RenderX

ambulate independently. Because the wrist-worn device infers activity from arm motion, reduced arm swing during ambulation could cause genuine walking to be registered as sedentary, contributing to the overestimation observed. We were unable to evaluate or adjust for this effect, and future studies should record gait parameters such as arm-swing amplitude and cadence to examine it directly.

The exploratory multivariable analysis of moderators of the bias was underpowered (N=51; model P=.12; R^2=0.15), and the absence of statistically significant moderators carries a high risk of type II error. Nonsignificance does not establish the absence of an effect: the adjusted estimate for heart failure versus ischemic heart disease (+17.4 minutes) had a wide 95% CI (−3.7 to 38.5 minutes) whose upper bound exceeds half of the overall 62-minute bias and is therefore potentially clinically meaningful. Consequently, our data do not establish that a single population-level correction is adequate across clinically important subgroups; subgroup-specific calibration, particularly by primary diagnosis, cannot be excluded and should be evaluated in adequately powered studies.

Cultural and Regional Factors

The direction and magnitude of the observed bias may be shaped by culturally specific patterns of daily activity that are common in Japan but less so in Western settings. Traditional floor-based sitting postures (eg, seiza and cross-legged sitting) and many low-intensity household and self-care activities are typically performed with the trunk relatively upright but with limited arm displacement. Because a wrist-worn device infers activity from arm motion, such postures may be misclassified as sedentary even when the participant is not sitting in the conventional sense, plausibly contributing to the overestimation we observed; conversely, validation studies in populations with more arm-dynamic routines have sometimes reported underestimation. Our cohort also differed from many North American and European outpatients with cardiac issues samples in its high mean age (78.4, SD 9 years) and substantial comorbidity burden, both of which tend to reduce ambulatory arm movement. These considerations indicate that neither the direction of bias nor the 62-minute correction factor should be assumed to transfer across cultural and regional contexts. For researchers and clinicians applying consumer wrist-worn devices in other settings, we recommend deriving a population-specific correction factor in a local validation sample, ideally against a posture-based reference such as the activPAL, before using device-derived sedentary time for quantitative purposes, and interpreting cross-population comparisons of absolute sedentary time with caution.

Comparator Considerations

We used the ActiveStyle Pro as the comparator, which, while validated against criterion measures (Douglas bag method) for energy expenditure, has not been compared directly with posture-based gold standards such as the activPAL for sedentary behavior classification. The activPAL is a thigh-worn inclinometer or accelerometer widely regarded as a reference method for classifying sitting or lying versus standing and for capturing postural transitions in adult field-based research [37-39].

Some residual measurement error in our comparator device cannot be excluded, which would tend to underestimate the true agreement between the Fitbit and criterion measures. Three-way validation studies comparing consumer devices, movement-based research accelerometers (ActiGraph and ActiveStyle Pro), and posture-based reference standards (activPAL) would strengthen the evidence base and clarify the relative performance of different monitoring approaches.

Our Fitbit sedentary classification relied on the metabolic-equivalent value returned with the Fitbit calories resource, which is itself an output of Fitbit's proprietary energy-expenditure algorithm. We necessarily assumed that this value corresponds to the same underlying construct as the metabolic-equivalent estimates produced by the ActiveStyle Pro, which were validated against the Douglas bag method. This assumption may not hold, because the two devices estimate metabolic equivalents using different and only partly disclosed algorithms. Part of the observed disagreement may therefore reflect differences between these algorithms, an unmeasured layer of algorithmic confounding, rather than the wrist-versus-waist placement difference alone, and we were unable to isolate this contribution with the available data.

Technology Evolution

Device software and algorithms are continuously updated by manufacturers, and our findings pertain specifically to the Fitbit Inspire 3 with the firmware version available during the study period (October 2022 to March 2023). Validation findings may not generalize across different Fitbit models (eg, Charge, Sense, and Versa series) or software versions. Manufacturers frequently update algorithms without notification, potentially altering device performance.

This highlights the need for ongoing surveillance and periodic revalidation as technology evolves. Establishing standardized validation protocols that can be rapidly deployed when new devices or firmware updates are released would help maintain evidence currency.

This study did not include an a priori sample size calculation; consecutive eligible patients enrolled during the study period were analyzed. The sample was therefore not designed to estimate intermethod agreement, and in particular the limits of agreement, with high precision, and the limits of agreement should be interpreted with corresponding caution.

Despite these limitations, this study provides clinically relevant evidence on the measurement characteristics of a widely used consumer-grade wearable in an understudied, high-risk cardiac population and offers a pragmatic foundation for future longitudinal validation and intervention studies in digital cardiovascular care.

Future Research Directions

Several avenues for future research emerge from this work.

Longitudinal validation studies should evaluate test-retest reliability, minimal detectable change, and responsiveness to behavioral interventions in cardiac populations. Studies should include repeated measurements over weeks to months to characterize within-person measurement error and determine

XSL•FO
RenderX

whether devices can reliably detect clinically meaningful changes.

Prospective intervention trials should assess whether Fitbit-detected changes in sedentary time correspond to changes measured by research-grade devices and whether these changes predict clinical outcomes (cardiovascular events, hospital readmissions, functional capacity, and quality of life). Such studies would establish the clinical utility of consumer devices for outcome assessment.

Machine learning and algorithm development studies could explore whether using raw Fitbit accelerometer data (rather than proprietary processed outputs) with population-specific classification algorithms can improve sedentary time accuracy in clinical populations. Recent advances in machine learning may enable personalized calibration approaches that account for individual movement signatures.

Multidevice validation studies would clarify relative performance and inform device selection for different research contexts. Such studies should incorporate a posture-based reference standard such as the activPAL, movement-based research accelerometers such as the ActiGraph and the ActiveStyle Pro, and multiple consumer devices such as the Fitbit, Apple Watch, and Garmin (Garmin Ltd).

Moderator analyses should systematically investigate sources of individual variability in measurement error, including patient characteristics (age, sex, body composition, cardiovascular functional capacity, and comorbidities), device factors (wear location and band tightness), and activity patterns (occupational demands and activity repertoire). Identifying moderators could enable stratified correction approaches or inform patient-device matching.

Pattern-based validation should assess device accuracy for sedentary bout duration, breaks in sedentary time, and temporal distribution patterns—metrics potentially more important than total volume for cardiovascular health.

Implementation research should examine patient and clinician perspectives on wearable-based sedentary behavior monitoring in real-world cardiac care settings, addressing questions of feasibility, acceptability, data interpretation, integration into clinical workflows, and impact on behavior change counseling effectiveness.

Cross-cultural validation should extend this work to diverse geographic and ethnic populations, recognizing that cultural practices (floor-sitting and traditional activities), built environments, and social norms around activity may influence device performance.

Implications for Clinical Practice and Research

These findings have important implications for both clinical practice and research in digital cardiovascular care. From a clinical perspective, the strong correlation and absence of proportional bias indicate that consumer-grade wrist-worn activity trackers can be used to monitor relative changes in sedentary behavior over time in outpatients with cardiac issues. This property is particularly relevant for remote monitoring, lifestyle counseling, and secondary prevention programs, where trends and behavioral patterns are often more informative than absolute values. However, clinicians should interpret absolute sedentary time estimates with caution, as wrist-worn devices may systematically overestimate sedentary behavior in older patients with cardiac issues. Applying a simple population-specific correction, as demonstrated in this study, may help contextualize tracker-derived metrics when setting behavioral targets or evaluating intervention effects.

From a research perspective, our results suggest that consumer-grade wrist-worn devices should not be used interchangeably with research-grade accelerometers for precise quantification of sedentary time without appropriate calibration. The moderate uncorrected concordance observed in this study indicates that such devices may be suboptimal as primary outcome measures when absolute sedentary time is required. Nevertheless, the substantial improvement in agreement after constant correction and the acceptable categorical agreement support their potential utility for large-scale observational studies, ranking-based analyses, and intervention trials focused on within-person change. Importantly, these findings underscore the necessity of population-specific validation and calibration when deploying consumer wearables in clinical research, particularly in older or functionally limited populations.

Conclusions

In older Japanese outpatients with cardiac issues, the Fitbit Inspire 3 systematically overestimated sedentary time by approximately 62 minutes within a standardized 12-hour daytime window compared with the waist-worn ActiveStyle Pro comparator, with significant fixed bias but no proportional bias. The between-device correlation was strong (r=0.905), and a simple constant correction improved concordance from moderate (CCC=0.71) to high; cross-validation indicated that this in-sample improvement was internally stable (out-of-sample CCC=0.90). In an exploratory analysis, the bias was not significantly moderated by sex, age, primary diagnosis, or musculoskeletal comorbidity, although this analysis was underpowered and a clinically relevant difference, particularly by primary diagnosis, cannot be excluded. Of greater practical importance, the wide limits of agreement indicate substantial interindividual variability that constrains the device's precision for individual-level measurement, and categorical agreement was only moderate (κ=0.59). These findings indicate that, in this population, a bias-corrected consumer wrist-worn tracker can provide group-level estimates and relative comparisons of sedentary time, but should not be treated as interchangeable with a research-grade comparator for precise individual quantification. Because the correction factor is specific to this population, device generation, and 12-hour measurement window, it should not be transferred to other settings without independent validation.

XSL•FO
RenderX

Acknowledgments

The authors used a generative AI assistant (Claude, Anthropic) during the revision process to support statistical reanalysis (leave-one-out cross-validation and exploratory subgroup regression), to verify the consistency of reported values against the analysis dataset, and to help draft and edit text. All analyses were independently verified by the authors against the source data, and the authors reviewed and take full responsibility for the final content of the manuscript.

Funding

This work was supported by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI; Grant Number JP23K10050).

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to ethical and privacy restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: KN, KPI, NM, MK, YK, HT, HI
Methodology: KN, KPI, NM, MK
Validation: KN, KPI, NM, MK
Formal analysis: KN, KPI, NM, MK
Investigation: YK, HT, HI
Supervision: KPI
Writing–original draft: KN
Writing–review & editing: KPI, NM, MK, YK, HT, HI

Conflicts of Interest

None declared.

References

1. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75. [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
2. Young DR, Hivert MF, Alhassan S, et al. Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American Heart Association. Circulation 2016 Sep 27;134(13):e262-e279. [doi: 10.1161/CIR.0000000000000440] [Medline: 27528691]
3. Onagbiye S, Guddemi A, Baruwa OJ, et al. Association of sedentary time with risk of cardiovascular diseases and cardiovascular mortality: a systematic review and meta-analysis of prospective cohort studies. Prev Med 2024 Feb;179:107812. [doi: 10.1016/j.ypmed.2023.107812] [Medline: 38081421]
4. Ajufo E, Kany S, Rämö JT, et al. Accelerometer-measured sedentary behavior and risk of future cardiovascular disease. J Am Coll Cardiol 2025 Feb 11;85(5):473-486. [doi: 10.1016/j.jacc.2024.10.065] [Medline: 39545903]
5. LaMonte MJ, LaCroix AZ, Nguyen S, et al. Accelerometer-measured physical activity, sedentary time, and heart failure risk in women aged 63 to 99 years. JAMA Cardiol 2024 Apr 1;9(4):336-345. [doi: 10.1001/jamacardio.2023.5692] [Medline: 38381446]
6. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462. [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
7. Naito K, Izawa KP, Maeda N, Kasai Y, Iwama H. Characteristics of physical activity and sedentary behavior in patients undergoing outpatient cardiac rehabilitation. Sci Rep 2024 Oct 16;14(1):24217. [doi: 10.1038/s41598-024-75362-9] [Medline: 39415044]
8. Izawa KP, Oka K. Focusing on sedentary behavior in comprehensive cardiac rehabilitation. Circ Rep 2024 Nov 8;6(11):481-488. [doi: 10.1253/circrep.CR-24-0098] [Medline: 39525298]
9. Izawa KP, Kitamura M, Ogura A, et al. Impact of sedentary time and physical function on health-related quality of life in hospitalized patients with cardiovascular disease. CJC Open 2025 Nov;7(11):1524-1532. [doi: 10.1016/j.cjco.2025.08.004] [Medline: 41425784]
10. van Bakel BMA, Kroesen SH, Bakker EA, et al. Effectiveness of an intervention to reduce sedentary behaviour as a personalised secondary prevention strategy for patients with coronary artery disease: main outcomes of the SIT LESS randomised clinical trial. Int J Behav Nutr Phys Act 2023 Feb 14;20(1):17. [doi: 10.1186/s12966-023-01419-z] [Medline: 36788615]

XSL•FO
RenderX

11. Patterson K, Davey R, Keegan R, et al. Testing the effect of a smartphone app on hospital admissions and sedentary behavior in cardiac rehabilitation participants: ToDo-CR randomized controlled trial. JMIR Mhealth Uhealth 2023 Oct 3;11:e48229. [doi: 10.2196/48229] [Medline: 37788043]
12. Saleh ZT, Aslanoğlu A, Almagharbeh WT, et al. Reducing sedentary behavior improves depressive symptoms among patients with heart failure enrolled in a home-based mobile health app cardiac rehabilitation. J Nurs Scholarsh 2025 May;57(3):394-403. [doi: 10.1111/jnu.13039] [Medline: 39663212]
13. Oshima Y, Kawaguchi K, Tanaka S, et al. Classifying household and locomotive activities using a triaxial accelerometer. Gait Posture 2010 Mar;31(3):370-374. [doi: 10.1016/j.gaitpost.2010.01.005] [Medline: 20138524]
14. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
15. Germini F, Noronha N, Borg Debono V, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791. [doi: 10.2196/30791] [Medline: 35060915]
16. Giurgiu M, Ketelhut S, Kubica C, et al. Assessment of 24-hour physical behaviour in adults via wearables: a systematic review of validation studies under laboratory conditions. Int J Behav Nutr Phys Act 2023 Jun 8;20(1):68. [doi: 10.1186/s12966-023-01473-7] [Medline: 37291598]
17. Hajj-Boutros G, Landry-Duval MA, Comtois AS, Gouspillou G, Karelis AD. Wrist-worn devices for the measurement of heart rate and energy expenditure: a validation study for the Apple Watch 6, Polar Vantage V and Fitbit Sense. Eur J Sport Sci 2023 Feb;23(2):165-177. [doi: 10.1080/17461391.2021.2023656] [Medline: 34957939]
18. Delobelle J, Lebuf E, Dyck DV, et al. Fitbit's accuracy to measure short bouts of stepping and sedentary behaviour: validation, sensitivity and specificity study. Digit Health 2024;10:20552076241262710. [doi: 10.1177/20552076241262710] [Medline: 38894943]
19. Rowlands AV, Olds TS, Hillsdon M, et al. Assessing sedentary behavior with the GENEActiv: introducing the sedentary sphere. Med Sci Sports Exerc 2014 Jun;46(6):1235-1247. [doi: 10.1249/MSS.0000000000000224] [Medline: 24263980]
20. Bailey DP, Ahmed I, Cooper DL, et al. Validity of a wrist-worn consumer-grade wearable for estimating energy expenditure, sedentary behaviour, and physical activity in manual wheelchair users with spinal cord injury. Disabil Rehabil Assist Technol 2025 Apr;20(3):708-714. [doi: 10.1080/17483107.2024.2405895] [Medline: 39301994]
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
22. Lin LIK. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989 Mar;45(1):255-268. [doi: 10.2307/2532051] [Medline: 2720055]
23. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
24. Collins JE, Yang HY, Trentadue TP, Gong Y, Losina E. Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. PLoS One 2019;14(1):e0211231. [doi: 10.1371/journal.pone.0211231] [Medline: 30699159]
25. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. J Med Internet Res 2016 Sep 7;18(9):e239. [doi: 10.2196/jmir.5531] [Medline: 27604226]
26. Carpenter C, Yang CH, West D. A comparison of sedentary behavior as measured by the Fitbit and ActivPAL in college students. Int J Environ Res Public Health 2021 Apr 8;18(8):3914. [doi: 10.3390/ijerph18083914] [Medline: 33917879]
27. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 7;19(1):29. [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
28. Salim A, Brakenridge CJ, Lekamlage DH, et al. Detection of sedentary time and bouts using consumer-grade wrist-worn devices: a hidden semi-Markov model. BMC Med Res Methodol 2024 Sep 30;24(1):222. [doi: 10.1186/s12874-024-02311-5] [Medline: 39350114]
29. Yano S, Koohsari MJ, Shibata A, Ishii K, Mavoa S, Oka K. Assessing physical activity and sedentary behavior under free-living conditions: comparison of active style pro HJA-350IT and ActiGraphTM GT3X. Int J Environ Res Public Health 2019 Aug 23;16(17):3065. [doi: 10.3390/ijerph16173065] [Medline: 31450754]
30. Kurita S, Yano S, Ishii K, et al. Comparability of activity monitors used in Asian and Western-country studies for assessing free-living sedentary behaviour. PLoS One 2017;12(10):e0186523. [doi: 10.1371/journal.pone.0186523] [Medline: 29045441]
31. Sasai H. Assessing sedentary behavior using wearable devices: an overview and future directions. J Phys Fit Sports Med 2017;6(3):135-143. [doi: 10.7600/jpfsm.6.135]
32. Wu J, Fu Y, Chen D, et al. Sedentary behavior patterns and the risk of non-communicable diseases and all-cause mortality: a systematic review and meta-analysis. Int J Nurs Stud 2023 Oct;146:104563. [doi: 10.1016/j.ijnurstu.2023.104563] [Medline: 37523952]
33. Swartz AM, Steinbrink GM, Strath SJ, Mitra T, Morelli WA. A systematic review of the effects of breaking up/interrupting sedentary behavior on health outcomes among community-dwelling adults 60+ years. J Aging Phys Act 2025 Jun 1;33(3):287-308. [doi: 10.1123/japa.2023-0416] [Medline: 39566490]

XSL•FO
RenderX

34. Liang YY, He Y, Huang P, et al. Accelerometer-measured physical activity, sedentary behavior, and incidence of macrovascular and microvascular events in individuals with type 2 diabetes mellitus and prediabetes. J Sport Health Sci 2025 Dec;14:100973. [doi: 10.1016/j.jshs.2024.100973] [Medline: 39214513]
35. Farrahi V, Collings PJ, Oussalah M. Deep learning of movement behavior profiles and their association with markers of cardiometabolic health. BMC Med Inform Decis Mak 2024 Mar 13;24(1):74. [doi: 10.1186/s12911-024-02474-7] [Medline: 38481262]
36. Hyde ET, Bandoli GE, Zou J, et al. Prospective associations between accelerometer-measured physical activity, sedentary behavior, and healthy longevity: the Women's Health Accelerometry Collaboration. J Gerontol A Biol Sci Med Sci 2025 Nov 6;80(12):glaf206. [doi: 10.1093/gerona/glaf206] [Medline: 41017648]
37. Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. Br J Sports Med 2006 Dec;40(12):992-997. [doi: 10.1136/bjsm.2006.030262] [Medline: 16980531]
38. O'Brien MW, Wu Y, Petterson JL, Bray NW, Kimmerly DS. Validity of the ActivPAL monitor to distinguish postures: a systematic review. Gait Posture 2022 May;94:107-113. [doi: 10.1016/j.gaitpost.2022.03.002]
39. Edwardson CL, Winkler EAH, Bodicoat DH, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178. [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]

Abbreviations

ACE: angiotensin-converting enzyme
ARB: angiotensin receptor blocker
ARNI: angiotensin receptor–neprilysin inhibitor
CAD: coronary artery disease
CCC: concordance correlation coefficient
CVD: cardiovascular disease
HCN: hyperpolarization-activated cyclic nucleotide
LOOCV: leave-one-out cross-validation
LPA: light-intensity physical activity
MET: metabolic equivalent
MVPA: moderate to vigorous physical activity
NYHA: New York Heart Association
RR: relative risk
SB: sedentary behavior
SGLT2: sodium-glucose cotransporter-2

Edited by A Coristine; submitted 14.Jan.2026; peer-reviewed by B Cao, Z Zhong; revised version received 17.Aug.2026; accepted 21.Aug.2026; published 30.Sep.2026.

Please cite as:
Naito K, Izawa KP, Maeda N, Kasai Y, Tani H, Kondo M, Iwama H
Assessment of Wrist-Worn Activity Tracker Accuracy for Measuring Sedentary Behavior in Outpatients With Cardiovascular Disease: Validation Study
JMIR Cardio 2026;10:e91104
URL: https://cardio.jmir.org/2026/1/e91104
doi:10.2196/91104

© Koichi Naito, Kazuhiro P Izawa, Noriaki Maeda, Yuya Kasai, Haruna Tani, Misaki Kondo, Hajime Iwama. Originally published in JMIR Cardio (https://cardio.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-Time Physical Activity Self-Monitoring in Patients Hospitalized for Cardiology Care: Interrupted Time Series Analysis

Carlos de Miguel Llorente[1], PT, MSci; Gertjan T J Sieswerda[2], MD, PhD; Kaoutar Karramass[1], PT, MSci; Charlotte Rommens[1], PT; Thomas J Hoogeboom[1,3], PT, PhD; Karin Valkenet[1,3], PT, PhD

[1]Department of Rehabilitation, Physiotherapy Science and Sport, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands
[2]Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands
[3]Group Innovation of Human Movement Care, University of Applied Sciences Utrecht, Utrecht, The Netherlands

Corresponding Author:
Carlos de Miguel Llorente, PT, MSci
Department of Rehabilitation, Physiotherapy Science and Sport, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands

Abstract

Background: Hospital admission is associated with increased sedentary behavior and low levels of physical activity. Hospitals have developed several strategies and interventions to address this unwanted inactivity and increase patient movement during admission. Self-monitoring of physical activity is a promising approach to support activity during hospital stays.

Objective: This study investigated whether providing patients with real-time physical activity feedback, compared to a context where only health care practitioners had access to activity data, supported patients in maintaining activity levels in the cardiology ward.

Methods: A Hybrid Type 2 interrupted time series design was applied. In phase 1 (24 wk), patients wore accelerometers (Physical Activity Monitor [PAM] AM400) with data visible only to health care professionals. In phase 2 (24 wk), self-monitoring was introduced using a ward-based screen that provided patients with real-time feedback on daily physical activity. Implementation outcomes were evaluated within the RE-AIM (Reach, Effectiveness, Adoption, and Implementation) framework, with "Maintenance," defined as daily physical activity trends over time, serving as the primary outcome. The other RE-AIM dimensions were assessed as secondary outcomes.

Results: A total of 159 patients were included (n=75 in phase 1 and n=84 in phase 2). Daily physical activity levels were expressed as active minutes per day. No significant immediate change in daily activity occurred at the start of phase 2 vs the end of phase 1 (β=–0.127, *P*=.81). In phase 1, physical activity declined statistically significantly over time (β=–0.002, *P*<.001; ~6% decrease per month). In phase 2, following the introduction of the self-monitoring intervention, this decline was no longer observed, and activity levels were maintained. A significant phase interaction (β=0.002, *P*=.03) confirmed stabilization of physical activity levels in phase 2. Secondary RE-AIM outcomes did not differ between phases.

Conclusions: The decline observed when only health care professionals accessed the data was no longer present once patients could monitor their own physical activity. Although seasonal influences cannot be excluded, these findings suggest that patient self-monitoring may support the maintenance of physical activity during hospital stays. Sustainability is complex, and determining the effect of patient self-monitoring alone remains challenging. Larger studies are needed to confirm these results.

(JMIR Cardio 2026;10:e92189) doi:10.2196/92189

KEYWORDS

inpatients; motor activity; self-monitoring; implementation science; cardiology

Introduction

Growing attention has been given to physical activity during hospital stays [1-4]. Inactivity is highly prevalent during hospitalization and is recognized by health care professionals as an important issue to be addressed [5-8]. Various interventions have been implemented in hospital settings to reduce this unwanted sedentary behavior, including strategies such as early mobilization and patient education [9-16].

Among these interventions, self-monitoring of physical activity is an effective strategy for improving patient activity during hospital stays [16-19]. By using movement monitors or accelerometers, clinicians can record patients' activity levels throughout admission. These data are valuable both for clinicians, as an outcome measure, and for patients, as feedback on their progress and recovery [18,20-22]. In a recent pilot study, hospitalized patients undergoing cardiac rehabilitation were randomly assigned either to standard activity monitoring

XSL•FO
RenderX

(with an accelerometer) or to an intervention combining activity monitoring with a physical activity dashboard and daily feedback [23]. The intervention group demonstrated significantly higher activity levels at discharge compared to the control group, showing the potential of physical activity monitoring during hospital stay.

However, once implemented, most interventions are rarely assessed to determine whether their effects are sustained over time [11,23-26]. Research in physical activity implementation has largely focused on effectiveness at specific time points using cross-sectional designs, providing limited insight into long-term sustainability [15,16]. However, sustainability is complex, and even perfectly delivered interventions may not thrive over time due to voltage drop (ie, the effectiveness diminishes in real-world conditions) or program drift (ie, the delivery gradually deviates from the intended protocol) [27]. Therefore, it is important to examine whether interventions designed to modify physical activity are maintained after their implementation. Implementation frameworks such as RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) are available to support such assessments, as they address not only how interventions are delivered but also the extent to which their use and effects are maintained over time [28]. In addition, analytical approaches such as interrupted time series (ITS) designs are particularly valuable in this context [29-31] as they allow for the assessment of both the immediate effects of an intervention and changes in trends over time.

Hence, this study investigated whether providing patients with real-time physical activity feedback, compared to receiving no real-time physical activity feedback, supported patients in maintaining physical activity levels in the cardiology ward. The evaluation was guided by the RE-AIM framework, with maintenance of the intervention as the primary outcome.

Methods

Study Design

This study used an ITS design using prospectively collected routine care data. A Hybrid Type 2 implementation-effectiveness approach was applied [32]. This allowed simultaneous evaluation of the self-monitoring intervention on daily physical activity and the implementation outcomes using the RE-AIM framework [32]. This study was reported in accordance with the SQUIRE (Standards for Quality Improvement Reporting Excellence) 2.0 [33] and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [34].

Intervention

The intervention consisted of the addition of a ward-based screen that provided patients with real-time feedback on their daily physical activity, introduced after a baseline phase as part of a 2-phase design.

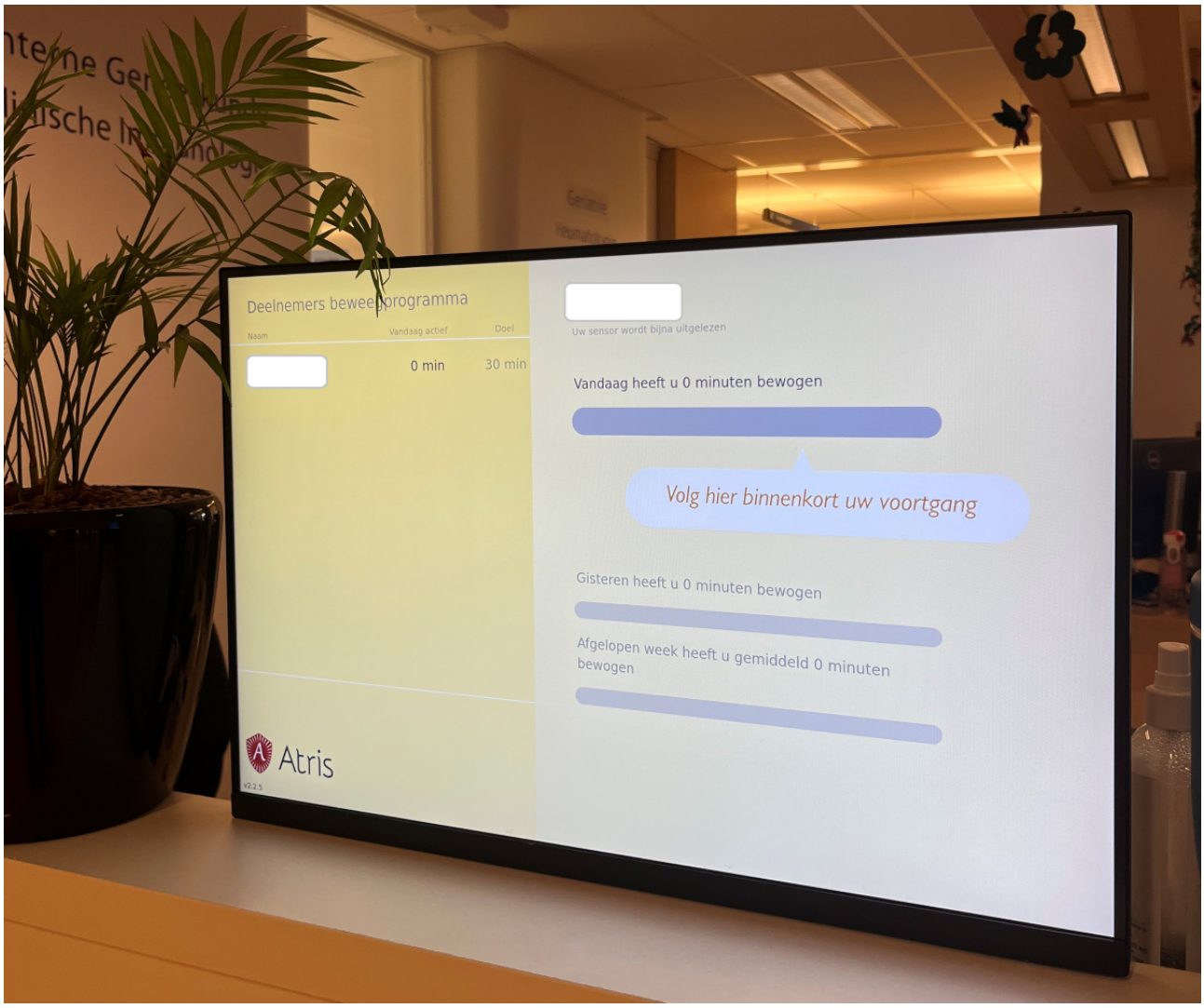
In phase 1 (April 2024 to September 2024). During this first phase, only the health care practitioners (HCPs) of the cardiology ward had access to the patient's daily physical activity data in the electronic health record (EHR).

In phase 2 (September 2024 to March 2025), a ward-based screen was added to supplement phase 1 activities, providing patients and HCPs with real-time insight into daily physical activity levels. Patients could view both their own daily physical activities and that of others in the ward. All patients were represented on the screen by an alias, ensuring that no real names were ever displayed. Patients could choose their own alias; if no preference was expressed, an alias was automatically assigned by the Atris system (Dawn Technology). The screen was positioned outside the patient rooms, directly in front of the ward's secretary desk. The location and setting of the screen are shown in Figure 1. Details on ward type and typical patient admissions, provided together with the implementation strategies used, are shown in Multimedia Appendix 1.

It is important to note that phase 1 did not represent a routine care baseline. During phase 1, HCPs already had access to patients' daily physical activity data via the EHR web viewer, and implementation strategies, including training of coaches and key users, and clinical lessons on the ward were already in place. The introduction of the ward-based screen in phase 2 was therefore the primary difference between phases. The comparison between phases reflected the effect of adding the ward screen to an already active implementation context, rather than a comparison between an intervention and a control period.

XSL•FO
RenderX

Figure 1. Example of the ward-based screen used for the self-monitoring intervention in the cardiology ward at the University Medical Center Utrecht (UMC Utrecht), the Netherlands, during phase 2 (September 2024 to March 2025). The screen, powered by the Atris platform, was positioned at the ward's secretary desk. The left panel displays a group overview of patients identified by an alias alongside their total daily physical activity in minutes. The right panel shows an individual patient view with today's, yesterday's, and weekly average activity. No patient names were displayed. The "Doel" (Goal) column was not activated during the study period.



Setting and Study Population

The self-monitoring intervention was implemented as part of routine care in the cardiology ward from the University Medical Centre of Utrecht (UMC), the Netherlands. Data were collected from patients and HCPs from the cardiology ward. Both patients and HCP had different data recorded and different inclusion and exclusion criteria. At the cardiology ward, patients with an expected hospital stay of at least 48 hours receive an accelerometer as part of routine care unless they are wheelchair-dependent. Patients were eligible for inclusion in the study when they were admitted to the cardiology ward and had at least 2 consecutive valid days of physical activity data recorded by the accelerometer. A valid day was considered any day that had at least 1 minute of total physical activity. The 1-minute threshold for a valid day was established based on consultation with HCPs and researchers with experience in the use of the Physical Activity Monitor (PAM) accelerometer in cardiology care. In regular clinical practice, patients recording 0 minutes of activity were found not to have worn the device at all, while any recording of at least 1 minute reliably indicated actual wear. This threshold therefore reflected a meaningful distinction between nonwear and wear days in this specific device and population context, rather than an arbitrary cutoff. Partial wear days, including admission and discharge days, were considered invalid measurement days, as variation in admission and discharge times could result in incomplete activity recordings that were not comparable across patients. Moreover, the requirement of at least 2 consecutive valid days was applied to ensure that the physical activity data represented a stable and meaningful period of the patient's hospital stay, rather than isolated or incomplete observations. Patients were excluded if they had objected against the use of their routine care data for research purposes or had a hospital stay shorter than 48 hours. HCPs were included if they were a doctor, nurse, or physiotherapist working in the cardiology ward, they used or were currently using the accelerometers in cardiology care, and they had direct clinician-patient interactions.

Ethical Considerations

The PAM accelerometers were implemented as part of routine care in the cardiology ward at the UMC Utrecht. Hence, this

XSL•FO
RenderX

study did not fall under the scope of the Dutch Medical Research Involving Human Subjects Act (WMO) and therefore did not require approval from an accredited medical ethics committee. However, an independent quality check was conducted at UMC Utrecht to ensure compliance with applicable legislation and regulations, including informed consent procedures, data management, privacy, and legal requirements. Non-WMO permission was granted following protocol review (24U-0064-CARMMI). Hence, informed consent for the use of routine care data was waived. All patient data, including EHR data and accelerometer activity data, were pseudonymized prior to analysis, and no traceable information was available to the research team at any point. HCPs provided verbal consent for the collection of net promoter scores (NPSs) under the CARMMI study framework. Their data were also pseudonymized, with no traceable information available. No compensation was provided to patients or HCPs.

Study Outcomes

Implementation outcomes were evaluated using the RE-AIM framework, which assesses Reach, Effectiveness, Adoption, Implementation, and Maintenance of the self-monitoring intervention [28]. Data for each RE-AIM dimension were collected across both phases of the implementation process (Table 1).

Table . Overview of the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework components and corresponding outcome measures[a].

RE-AIM model component	Outcome
Reach	Percentage of eligible patients that received an accelerometer
Effectiveness	Average daily physical activity of the patients during the implementation phases
Adoption	Net promoter score results
Implementation	Percentage of wearing time per patient
Maintenance	Trends of daily physical activity within each of the implementation phases

[a]These were used to evaluate a ward-based self-monitoring intervention for physical activity in inpatients receiving cardiology care at the University Medical Centre Utrecht (UMC Utrecht). Outcomes were assessed across 5 RE-AIM dimensions: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

Reach was defined as the percentage of eligible patients admitted to the ward who were given an accelerometer. This was expressed as the percentage of patients who received the accelerometer divided by the total number of patients who were admitted to the ward during that phase. Eligible patients were those with a length of stay exceeding 48 hours who did not use wheelchairs. We prespecified ≥70% as the implementation target [16].

Effectiveness, defined as the impact of the intervention, was assessed by comparing average daily physical activity levels between phases, summarized with geometric means (GMs) due to skewed data. Physical activity data were measured with the PAM version AM400. The PAM is a 3-axial accelerometer, which is worn around the ankle and measures the amount of total activity time. Agreement with the well-established accelerometer Actigraph (wGT3X-Bt) showed a strong intraclass correlation coefficient of 0.955 in patients from different wards [35]. The PAM registers movements above the threshold of 1.4 metabolic equivalent task (MET), which are labeled as active minutes. The PAM provides a daily value of physical activity expressed in active minutes per day (eg, 17 minutes/day). This output was expressed as daily physical activity levels. HCPs did not perform any physical activity data entry or collection. Upon patient admission, ward staff connected the PAM, after which data collection began automatically. No further action was required from HCPs or patients beyond ensuring continued device wear. In addition, the following demographic and clinical characteristics of the patients were obtained to be used as covariates for the analysis: age (years), sex (male/female), height (cm), weight (kg), and BMI (kg/m^2). Length of stay, discharge destination, and reason of admission were also retrieved from the EHR. To classify the reason for admission, admission diagnoses were categorized into clinical subgroups (eg, acute coronary syndromes, heart failure, cardiac rhythm disorders, and pericardial infection/inflammation) in collaboration with HCPs from the cardiology ward.

Adoption was defined within the RE-AIM framework as the willingness of HCPs to deliver the intervention. Therefore, HCPs were invited to complete a brief survey incorporating the NPS methodology [36]. The NPS survey is simple and widely used, but research in health care suggests its validity is modest and caution is needed when interpreting single-item scores [36,37]. HCPs were approached in person, and verbal consent was obtained prior to participation. The NPS survey was administered once during phase 1 to establish baseline perceptions and twice during phase 2 to monitor changes in HCPs' experience over time. As the surveys were anonymous and completed by different HCPs on each occasion, these represented 3 independent cross-sectional measurements rather than a repeated-measures design. The survey consisted of 2 questions rated from 0 to 10: "On what level do you recommend the PAM to other colleagues?" and "On what level do you recommend the PAM to patients?" The surveys were completely anonymous, and different HCPs answered these surveys each time. Although the NPS is traditionally categorized into promoters (9-10), passives (7-8), and detractors (0 - 6), with the final score calculated as the percentage of promoters minus detractors, this approach was not used in the present study. Instead, NPS responses were analyzed as a continuous variable on the original 0 to 10 scale, an approach commonly adopted

XSL•FO
RenderX

for smaller samples [36]. Scores were summarized using means and SDs.

Implementation was defined as the extent to which the intervention was delivered as intended, that is, whether patients wore the accelerometer every day during their hospital admission. Therefore, implementation was expressed as the wearing time (percentage of wearing days per patient) in each of the phases. Because no consensus threshold exists for accelerometer adherence in hospitalized populations, we arbitrarily considered ≥60% of admission days as satisfactory. Therefore, the proportion of patients meeting this threshold in each phase was also reported.

Maintenance of the intervention, the primary outcome of the study, was evaluated as the trend in daily physical activity levels in the ward within each phase. This outcome therefore reflected the sustainability of the intervention by assessing whether average ward-level activity was maintained across phases and whether there were differences between them.

Data Analyses

All analyses were conducted in Python (Python Software Foundation; version 3.11.9) using stats models, pandas, and SciPy. Days without accelerometer wear time were excluded from the analysis. For patients with admissions in both phase 1 and phase 2, only data from their first admission were included to minimize bias from prior exposure.

Baseline demographic and clinical characteristics were summarized separately for patients in phase 1 and phase 2. Continuous variables (ie, age, height, weight, BMI, length of stay, and wearing time) were reported as means and SDs. Group differences were tested using the Mann-Whitney *U* test. Categorical variables (ie, sex, discharge destination, and reason for admission) were reported as counts and percentages and compared across phases using chi-squared test.

Our primary outcome for the analysis was the RE-AIM component Maintenance. Because physical activity was log-normally distributed, we log-transformed values and fit an ITS using a linear mixed effects model with restricted maximum likelihood estimation to estimate within-phase trends and the immediate change at phase 2 onset (immediate change in physical activity at the beginning of phase 2) [31,38,39]. Fixed effects included time (days since the start of observation), phase (phase 1 vs phase 2), and their interaction. A random intercept was added for each patient to account for repeated measures. In this specification, the time coefficient represents the trend in phase 1. The phase coefficient captures the immediate level change at the start of phase 2, and the interaction term quantifies any change in trend following the intervention.

The secondary outcomes (the remaining RE-AIM components) were compared between the 2 phases. First, differences in Reach between phases were compared using a 2-proportion z-test. Second, to evaluate the Effectiveness outcome, we calculated the average treatment effect of the intervention on physical activity with a weighted least squares regression model. This model included inverse probability of treatment weighting (IPW) with propensity scores to adjust for baseline confounding. Propensity scores were calculated via logistic regression using baseline covariates. Weights were calculated as the inverse of the estimated probability of receiving the actual treatment. Standardized mean differences were calculated for all covariates before and after weighting to evaluate covariate balance. Physical activity, the dependent variable, was log-transformed, so model coefficients could be interpreted as geometric mean ratios.

Third, differences in Adoption, measured by NPS across phase 1, phase 2 (beginning), and phase 2 (end), were compared using a 1-way ANOVA, as normality assumptions were met. Fourth and last, Implementation between phases was evaluated using the Mann-Whitney *U* test due to nonnormal distribution. The proportion of patients meeting the threshold was compared using the chi-square test. A significance level of α=.05 was used for all statistical tests.

Reflexivity

As researchers focused on physical activity during hospitalization, we acknowledge that we anticipated the addition of real-time physical activity self-monitoring would help maintain, and possibly increase, patient activity levels. To reduce the influence of these assumptions, all analyses were conducted using predefined procedures and independently reviewed by a statistician to ensure accurate interpretation. Although some authors (CR, KK, and GTJS) were involved in implementing the 2-phase intervention within the ward, the authors who led the design, interpretation, and writing of this paper (CdML, KV, and TJH) were not involved in daily cardiology ward procedures and did not influence clinical practice during the study period. Findings were discussed with the multidisciplinary team directly involved in implementation to ensure that multiple perspectives were considered.

Results

Baseline Characteristics

In total, data were recorded for 233 hospital admissions across the 2 phases. Of these, 9 patients had admissions in both phases, of whom 5 had three admissions across the study period, accounting for 14 duplicate admissions in total. Only the first admission per patient was retained, resulting in 219 unique patient admissions. Of these, 60 patients were excluded due to having fewer than 2 days of recorded daily physical activity, leading to a final sample of 159 patients included in the study. No further information was available regarding the reasons for discontinuous activity recordings or the absence of valid measurement days. In total, 75 patients had their daily physical activity recorded during phase 1 and 84 patients during phase 2 (Table 2). No significant differences were found regarding age (*P*=.15) and length of stay (*P*=.73) between patients from both phases. Patients in phase 2 had significantly lower BMI compared to patients in phase 1 (*P*=.01).

XSL•FO
RenderX

Table . Baseline characteristics of cardiology inpatients at the University Medical Centre Utrecht (UMC Utrecht) across 2 implementation phases: phase 1 (April 2024 to September 2024, n=75) and phase 2 (September 2024 to March 2025, n=84)[a].

Outcome (statistic)	Phase 1 (n=75)	Phase 2 (n=84)	*P* value
Age, median (IQR)	66 (21 - 88)	59 (18 - 89)	.15
Weight, median (IQR)	82.5 (42.9 - 169)	80 (44.3 - 135)	.17
Height, median (IQR)	176 (140 - 200)	177.5 (155-207)	.24
BMI, median (IQR)	26.9 (20.1 - 43.6)	24.3 (16 - 33.7)	.01[b]
Length of stay, median (IQR)	12 (2 - 75)	12 (4-62)	.73
Sex, n (%)			.64
Female	24 (32)	23 (27.4)	
Male	51 (68)	61 (72.6)	
Reason for admission, n (%)			.75
Observations and general admission	8 (10.7)	9 (10.7)	
Cardiac arrest/resuscitation	5 (6.7)	10 (11.9)	
Pericardium infection/inflammation	5 (6.7)	5 (6)	
Myocardial ischemia and infarction	24 (32)	18 (21)	
Heart failure and myocardial insufficiency	21 (28)	25 (29.8)	
Valve abnormalities	2 (2.7)	2 (2.4)	
Rhythm and conduction disorders	10 (13.2)	15 (17.9)	
Discharge destination, n (%)			.64
Home	60 (80)	73 (86.9)	
Other hospital	6 (8)	6 (7.1)	
Nursing home	1 (1.3)	0 (0)	
Rehabilitation hospital	6 (8)	4 (4.8)	
Died in hospital	2 (2.7)	1 (1.2)	

[a]Values are presented as median (IQR) for continuous variables and n (%) for categorical variables. Between-phase differences were assessed using the Mann-Whitney *U* test for continuous variables and the chi-square test or Fisher exact test for categorical variables.

[b]$P<.05$ was considered statistically significant.

Primary Outcome: Maintenance

The trend analysis of the restricted maximum likelihood model showed that daily physical activity levels of patients significantly declined during phase 1 (β=–0.002, $P<.001$), corresponding to an estimated 5.9% decrease of daily physical activity per month (exp (–0.002×365.25/12) ≈ 0.941).

During phase 2, the trend analysis showed that daily physical activity stabilized, and the trend was not significantly different from 0 (β=0.000). The time trends significantly differed between phases, with a positive interaction term (β=0.002, $P=.03$), indicating a flattening of the downward trend of daily physical activity in phase 2 (Table 3). The results of the physical activity trends per phase are shown in Figure 2. No statistically significant immediate change in daily physical activity levels was observed at the start of phase 2 right after the end of phase 1 (β=–0.127, $P=.81$). In addition, to account for the significant difference in baseline BMI between phases ($P=.01$), the primary model was additionally adjusted for BMI. The interaction term remained statistically significant (β=0.002, $P=.02$).

XSL•FO
RenderX

Table . Results of the linear mixed effects interrupted time series model for daily physical activity (log-transformed and non-transformed) across implementation phases among inpatients in cardiology care at the University Medical Centre Utrecht (UMC Utrecht) (phase 1: April 2024 to September 2024, n=75; phase 2: September 2024 to March 2025, n=84)[a].

Fixed effects	Log-transformed total activity β[b] (SE)	*z* [c]	*P* value	Total activity β (SE)	*z*	*P* value	Log-transformed (BMI-adjusted[d]) β (SE)	*z*	*P* value
Intercept	3.710 (0.185)	20.025	<.001	54.442 (6.372)	8.544	<.001	4.309 (0.310)	13.919	<.001
Time index	–0.002 (0.001)	–3.720	<.001	–0.050 (0.019)	–2.572	.01	–0.002 (0.001)	–3.787	<.001
Phase 2	–0.127 (0.532)	–0.239	.81	–26.547 (18.570)	–1.430	.15	–0.250 (0.529)	–0.472	.64
Time × Phase 2	0.002 (0.001)	2.212	.03	0.071 (0.028)	2.545	.01	0.002 (0.001)	2.381	.02
BMI	—[e]	—	—	—	—	—	–0.022 (0.009)	–2.430	.02

[a]The time index reflects the within-phase daily trend, phase 2 reflects the level change at the introduction of the ward-based screen, and the Time × Phase 2 interaction reflects the difference in slope between phases.

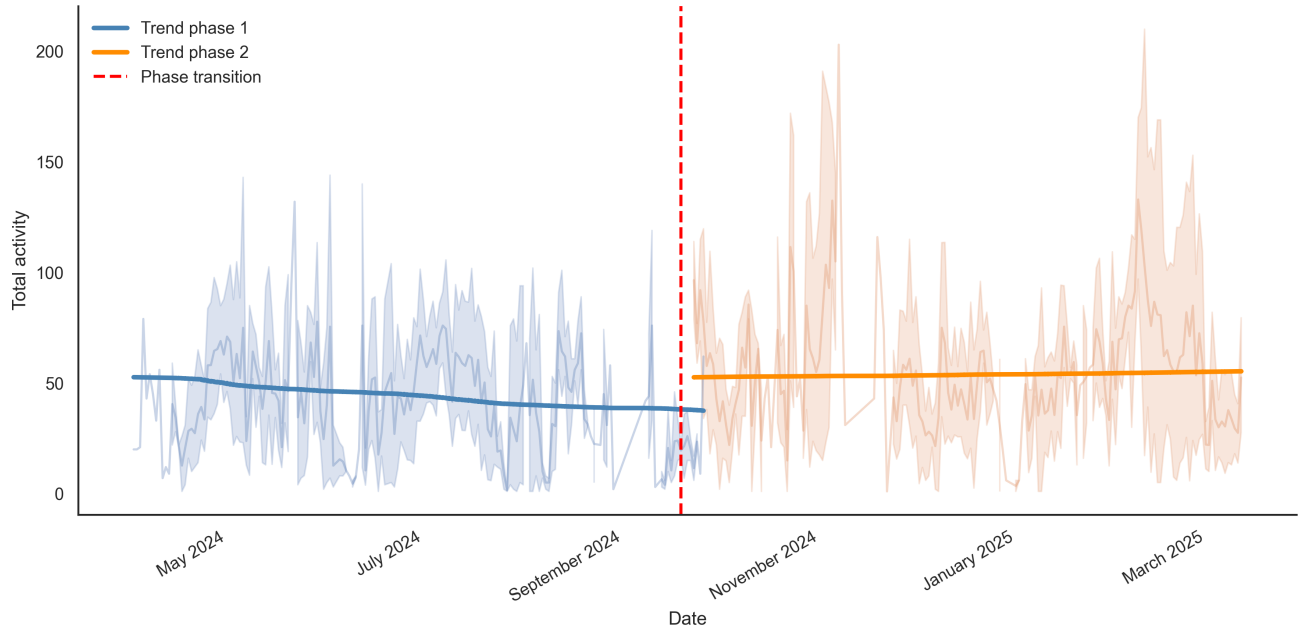
[b]β: regression coefficient.

[c]*z*: *z* statistic.

[d]The BMI-adjusted model includes BMI as a covariate.

[e]Not applicable.

Figure 2. Trends in daily physical activity (min) across the 2 implementation phases of a ward-based physical activity self-monitoring intervention in cardiology inpatients at the University Medical Center Utrecht (UMC Utrecht), the Netherlands. Phase 1 (April 2024 to September 2024, n=75): health care practitioners had access to patient activity data via the electronic health record only. Phase 2 (September 2024 to March 2025, n=84): a ward-based screen additionally provided patients with real-time insight into daily physical activity. The y-axis displays individual patient-day observations; shaded areas represent the full distribution of observed values. Solid lines represent separate linear trends fitted for each phase. The vertical dashed red line marks the transition from phase 1 to phase 2. A declining trend was observed during phase 1 (β=–0.002, *P*<.001), followed by stabilization in phase 2 (β=0.000, *P*>.05).



Three sensitivity analyses were performed to assess the robustness of the primary outcome. First, to assess the potential influence of seasonality, a sensitivity analysis was conducted incorporating calendar month as a covariate in the primary model. Calendar month was not significantly associated with daily physical activity levels (β=0.001, *P*=.98), and the interaction term remained in the same direction (β=0.002). Second, to assess the influence of extreme activity values on the primary outcome, a sensitivity analysis was conducted excluding individual measurements exceeding 200 activity units (n=4 observations, 0.4% of the data, from 2 patients). The results were consistent with the primary analysis: daily physical activity significantly declined during phase 1 (β=–0.002, *P*<.001), stabilized during phase 2 (net slope≈0.000), and the interaction

XSL•FO
RenderX

term remained statistically significant (β=0.002, *P*=.02). Third and last, to assess the potential influence of patients with extended length of stay on the observed activity trajectories, a sensitivity analysis was conducted excluding patients with a length of stay exceeding 30 days (n=14 patients, 193 observations). The results remained consistent with the primary analysis: daily physical activity significantly declined during phase 1 (β=–0.002, *P*<.001), stabilized during phase 2 (net slope≈0.000), and the interaction term remained statistically significant (β=0.002, *P*=.04).

The results of the ITS primary analysis are presented in Table 3. The results based on nontransformed daily physical activity were consistent with the log-transformed model, supporting the robustness of the findings. Observed total activity measurements and corresponding fitted values over time are shown in Multimedia Appendix 2.

Secondary Outcomes

For Reach, in phase 1, 36.6% (105/287 who were eligible based on length of stay) of admitted patients received an accelerometer, compared to 36.2% (128/354) in phase 2. This difference was not statistically significant (*P*=.91).

For Effectiveness, the GM of daily physical activity was 20.5 minutes for phase 1 and 23.5 minutes for phase 2. The exponentiated coefficient of the treatment effect (β=0.132) yielded a geometric mean ratio of 1.141, indicating a 14.1% higher average daily physical activity among patients in phase 2 compared to those in phase 1. However, this difference was not statistically significant (*P*=.52, 95% CI 0.761 - 1.709). The results of the IPW are shown in Multimedia Appendix 3.

For Adoption, mean recommendation scores to colleagues were 7.1 (SD 1.30) in phase 1 (n=17), 7.2 (SD 1.40) in the beginning of phase 2 (n=10), and 6.8 (SD 1.22) at the end of phase 2 (n=19). One-way ANOVA indicated no statistically significant differences in recommendation to colleagues between phases ($F_{2,43}$=0.332, *P*=.72). The mean recommendation scores to patients were 7.4 (SD 1.3) in phase 1, 7.5 (SD 1.7) in the beginning of phase 2, and 6.5 (SD 1.1) at the end of phase 2. These differences were also not statistically significant ($F_{2,43}$=2.900, *P*=.07).

Fourth and final, for Implementation, the median wearing time in phase 1 was 60.0% (IQR 42.0% - 74.5%) compared to 66.7% (IQR 50.0% - 78.9%) in phase 2; this difference was not statistically significant (Mann-Whitney *U*=3563.5, *P*=.15). Regarding the predefined implementation threshold of ≥60% wearing time, 52 of 75 (69.3%) patients in phase 1 and 52 of 84 (61.9%) patients in phase 2 met this threshold, with no statistically significant difference between phases (χ^2_1=0.666, *P*=.41). The results of the secondary analyses are displayed in Table 4.

XSL•FO
RenderX

Table . Summary of implementation outcomes evaluated using the RE-AIM (Reach, Effectiveness, Adoption, and Implementation) framework across 2 phases of a ward-based physical activity self-monitoring intervention in cardiology inpatients at the University Medical Centre Utrecht (UMC Utrecht), the Netherlands (phase 1: April 2024 to September 2024, n=75; phase 2: September 2024 to March 2025, n=84)[a].

RE-AIM[b] dimension and outcome	Phase 1	Phase 2	*P* value
Reach			
Eligible patients, %	36.6	36.2	.91
Effectiveness			
Changes in average daily physical activity levels (min)	20.5	23.5	.52
Adoption, mean (SD)			
NPS[c] colleagues	7.1 (1.3)	7.2 (1.4) - 6.8 (1.2)[d]	.72
NPS patients	7.4 (1.3)	7.5 (1.7) - 6.5 (1.1)[d]	.07
Implementation, median (IQR)			
Percentage of wearing time	60.0 (42.0 - 74.5)	66.7 (50.0 - 78.9)	.15
Maintenance			
Trends of daily physical activity			.03[e]
Trend coefficient (β)	–0.002	0.000	
Activity change per month, %	–5.9	0.0	

[a]Reach was defined as the percentage of eligible patients who received an accelerometer. Effectiveness reflects the geometric mean of daily physical activity in minutes, derived from a weighted least squares model on log-transformed values. Adoption was measured using the net promoter score (NPS) methodology, reported as mean (SD) for recommendations to colleagues and patients separately, and compared across 3 time points using one-way ANOVA. Implementation reflects the percentage of wearing time per patient, reported as median (IQR) and compared using the Mann-Whitney *U* test. Maintenance reflects trends in daily physical activity.

[b]RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

[c]NPS: net promoter score.

[d]Phase 2 values: beginning (n=10) and end (n=19).

[e]*P*<.05 is significant.

Discussion

Sumary of Outline

A ward-based screen that provided patients with real-time physical activity feedback was introduced after a baseline phase as part of a 2-phase design. In both phases, patients at the cardiology ward wore an accelerometer. The key distinction in phase 2 was the addition of visual feedback via a ward-based screen, which allowed patients to visualize both their own daily physical activity and that of other patients.

Principal Findings

The results of the primary outcome, Maintenance, showed that daily physical activity levels of the ward declined statistically significantly over time in phase 1. During this first phase, patients did not have access to their daily physical activity data unless an HCP explicitly shared it with them. However, accessing this information required HCPs to navigate to a specific setting in the EHR, which was likely not done consistently. This suggests that simply providing HCPs with access to patients' daily physical activity data via the EHR was possibly insufficient to sustain physical activity in the ward. In contrast, in phase 2, the negative time trend of daily physical activity was no longer observed. The addition of the ward-based screen where patients could visualize their physical activity was associated with a stabilization of physical activity levels. The addition of the ward-based screen possibly led to maintained levels of daily physical activity throughout phase 2, which lasted almost 6 months (24 wk). A possible explanation for this finding is that the introduction of the screen in the ward improved patients' intrinsic motivation. Providing patients access to their daily activity levels supports self-monitoring, a known behavior-change technique that increases awareness of mobility [10,15,40,41].

Regarding the other RE-AIM framework outcomes, no statistically significant differences were found. First, in terms of Reach, only slightly more than one-third of patients with an expected hospital stay of at least 2 days received an accelerometer. This proportion did not meet the predefined implementation threshold and is considered a major limitation in the implementation of an intervention [42]. The reasons for the low Reach remain unclear, as the research team was not involved in the process of providing accelerometers to patients or coordinating their allocation. It is possible that accelerometer distribution was influenced by factors such as nursing workload, shift variability, or patients' clinical status at the time of admission, which may have introduced selection bias. For example, patients who were less mobile, more acutely ill, or had a greater comorbidity burden may have been less likely to receive an accelerometer. Hence, the analyzed cohort may

XSL•FO
RenderX

overrepresent patients with better functional status, potentially leading to an overestimation of ward-level physical activity.

Second, under the Effectiveness outcome, the 14.1% increase in daily physical activity from phase 1 to phase 2 was not statistically significant and may be due to chance. Still, the change could be clinically relevant. As previously mentioned, in hospitalized patients, even modest increases in daily physical activity have been associated with improved functional outcomes, shorter lengths of stay, and reduced risk of readmission [43,44]. For instance, Gallardo-Gómez et al [45] conducted a systematic review and meta-analysis evaluating the effects of various daily physical activity interventions on functional outcomes during hospitalization. They identified a minimal effective dose of daily physical activity during hospitalization at approximately 100 METs-minutes/day, which corresponds to 40 minutes of light activity or 25 minutes of moderate activity per day. In our study, although direct comparisons are limited due to differences in measurement methods (accelerometer vs MET-based estimation), the GM of physical activity during phase 2 reached approximately 23.5 minutes per day. This level closely approaches the 25-minute/day threshold identified in the literature as the minimal effective dose to reduce functional decline. Therefore, it is possible that a portion of patients in phase 2 reached or exceeded this threshold, and that the observed increase in daily physical activity may have contributed to a reduced risk of functional deterioration in these patients. In addition, it is important to note that this study did not include a power calculation, as the sample size was determined by the available data from the implementation period. Hence, the study may have been underpowered to detect small but clinically meaningful differences in daily physical activity between phases, and the nonsignificant effectiveness finding should be interpreted with this limitation in mind.

Third, regarding Adoption, NPS scores were similar across phases (around 7/10), indicating no improvement in HCP engagement. Scores of this level are considered modest and suggest limited support or promotion of the intervention by HCPs. While NPS is a quick and accessible tool for assessing stakeholder (HCPs) support, its simplicity may limit its ability to capture more subtle aspects of HCP engagement [36]. It is possible that changes in engagement occurred but were not detected due to the tool's limitations. More detailed or sensitive evaluation methods may have offered deeper insight into HCP adoption. On the other hand, as HCPs also had direct insight into the patients' daily physical activity after the addition of the screen, this could have enabled HCPs to encourage physical activity more actively, thereby increasing patients' extrinsic motivation. Additionally, the visibility of other patients being active may have started a positive behavioral cascade within the cardiology ward [42,46,47]. As patients observed others engaging in physical activity, they may have felt inspired to do the same, reinforcing a culture of mobility and activity.

HCP engagement was considered closely related to the fourth and final RE-AIM outcome, Implementation. Implementation was measured as adherence. In phase 1, the mean wearing time was slightly below the predefined implementation threshold, whereas in phase 2 the mean exceeded the 60% threshold, indicating some improvement. We hypothesized that the screen would improve adherence by serving as a visual reminder, leading to HCPs to encourage patients to wear the device and increasing awareness when patients were not wearing it. However, results for wearing time remained stable across phases, with slightly higher percentages of wearing days in phase 2 compared to phase 1. While accelerometer-based self-monitoring has been shown to increase adherence outside the hospital setting [6,48], this effect was not clearly observed in our study. This may be related to the operational structure of the ward, particularly regarding accelerometer (PAM) ownership and responsibility. In the cardiology ward, patients with short hospital stays (less than 48 h) did not receive an accelerometer, meaning not all patients on the ward wore one. It is possible that HCPs did not consistently notice when data were missing or did not feel responsible for ensuring that patients were wearing the device.

Comparison to Previous Work

It is important to recognize that the goal of feedback-based monitoring is not only to increase daily physical activity, but also to promote independent mobility [19,48,49]. A systematic review from Compernolle et al [40] showed that the use of objective self-monitoring tools, such as accelerometers, may improve adherence and support progress toward behavior change goals through awareness of inactivity. However, changes in patient inactivity require intentional application of behavior change techniques (BCTs). Interventions using activity trackers during hospital care have shown greater effectiveness in increasing physical activity levels compared to routine care [41]. However, as shown by Leeuwerk et al [10], the effectiveness of such interventions is closely linked to the type and number of BCTs used. Evidence suggests that interventions grounded in behavioral theory, which incorporate multiple BCTs and are paired with professional coaching, are more likely to produce meaningful increases in daily physical activity levels. In our case, the intervention involved solely the addition of a ward-based screen to provide real-time physical activity self-monitoring. This may explain the limited changes observed in the trends of phase 2. It is also possible that the addition of the ward-based screen may also have engaged other BCTs not originally anticipated, including peer comparison, self-monitoring, and intrinsic goal setting.

To date, there are no studies that evaluate changes in daily physical activity during hospital stays using an ITS design. Most ITS studies have been conducted in community settings, aiming to improve physical activity and reduce sedentary behavior in specific communities or patient groups [29,30,50,51]. Although pre-post analysis is a commonly used statistical approach, it does not capture how effects of an intervention on physical activity are maintained over time. ITS analysis was more suitable in this context, particularly for a Hybrid Type 2 design, where the aim was to simultaneously assess both the effect of the intervention and its implementation. In this study, ITS allowed us to model average changes and trends in daily physical activity across both phases and to estimate changes attributable to the intervention.

XSL•FO
RenderX

Strengths and Limitations

This study had several strengths. First, physical activity was objectively measured using accelerometers, reducing self-report bias. Second, the study was conducted in a real-world clinical environment, enhancing its generalizability. Third and last, multidisciplinary involvement in the design and interpretation of the study strengthened the fidelity of the implementation and the interpretation of the findings.

First, data from routine care often involve a high proportion of missing data. While this reflects the realities of clinical practice, it presents certain limitations. We chose to exclude nonwearing days from the analyses, whereas current research increasingly supports the use of imputation methods, particularly when the proportion of missing data is not substantial. In our study, 60 (27.4%) patients were excluded for not meeting the minimum wear criterion. This exclusion occurred prior to analysis and reflects an inclusion criterion rather than missing data within the included sample. Among the 159 patients included in the final analysis, completeness was high. As missingness was limited, we used complete-case analysis instead of multiple imputation, assuming data were missing at random [52]. We therefore reported analyses based on observed values, which reflect the real conditions during hospitalization. Second, it would have been valuable to conduct qualitative interviews to gain deeper insights into HCPs' experiences during the implementation process. However, this was not within the scope of the study, which was designed to focus strictly on RE-AIM outcomes and the quantitative data available. As a result, our evaluation was limited to the data that was collected. Third, the study did not include a comparable ward as a control group. This would have allowed for the comparison of outcomes across different settings. It could also have provided insight into implementation among different patient populations. Fourth, a potential limitation of this study is that the observed outcomes could partly reflect seasonal or temporal factors. These could include unmeasured secular trends, such as changes in ward protocols or other unmeasured co-interventions occurring during the study period, rather than the intervention itself. Although IPW was used to account for measured baseline differences between phases, it does not fully address temporal confounding arising from the sequential nature of phase 1 and phase 2. The ITS analysis and the sensitivity analysis incorporating calendar month as a covariate partially mitigated this limitation by explicitly modeling the underlying activity trend over time, independently of phase assignment.

Finally, several potential confounders should be acknowledged. Cognitive function was not included as an exclusion criterion, and patients with cognitive impairment may have been less able to recognize and interpret their own activity data, potentially reducing the intervention effect. In addition, differences in disease severity may have influenced both baseline activity levels and the capacity to respond to the intervention. Disease severity may also have affected activity restrictions, such as prescribed bed rest or limitations on mobility. However, these factors were not systematically recorded in the EHR and therefore could not be accounted for in the analysis.

Future Directions

Future studies could explore incorporating additional BCTs to strengthen the self-monitoring intervention and further support physical activity maintenance during hospitalization. Additionally, future research should consider including information on the functional status and physical activity levels of patients prior to admission, as these factors will likely determine their activity upon hospitalization. These studies should move beyond recording total physical activity levels to also capture inactivity patterns, including sedentary bouts and moments of inconsistent activity. The relative effects of prolonged sedentary behavior and different patterns of physical activity accumulation remain poorly understood in hospitalized patients. For example, it is unclear whether 20 consecutive minutes of activity provides similar benefits to the same amount of activity accumulated through shorter bouts distributed throughout the day. Finally, future studies should examine within-day patterns of physical activity during hospitalization to better understand how real-time feedback on activity and inactivity, rather than daily activity totals alone, influences patient behavior and physical activity levels.

Conclusions

This study investigated whether providing patients with real-time physical activity feedback supported the maintenance of activity levels in the cardiology ward compared with receiving no feedback. Daily ward physical activity levels were maintained following the introduction of the ward-based screen, suggesting that real-time access to activity data may support physical activity maintenance during hospitalization.

Acknowledgments

The authors would like to thank the health care professionals of the cardiology ward for their dedicated work, as well as the implementation coaches and key users for their valuable support throughout the implementation process. This manuscript was supported by the use of ChatGPT-4 (OpenAI) to rephrase sentences, check grammar and vocabulary, and improve overall readability. The authors carefully reviewed and edited all AI-generated content to ensure accuracy and integrity of the final text.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

XSL•FO
RenderX

Data Availability

The data and full code, including preprocessing steps, filtering logic, variable derivation, and statistical analyses that support the findings of this study are available from the corresponding author upon reasonable request. Access to raw patient data is restricted due to ethical and privacy considerations.

Authors' Contributions

Conceptualization: CdML, KV
Formal analysis: CdML
Implementation: GTJS, KK, CR
Methodology: CdML, KV
Project administration: CdML, KV, TJH
Resources: GTJS
Writing – original draft: CdML
Writing – review and editing: CdML, KV, TJH, GTJS, KK, CR
All authors approved the final version of the manuscript.

Conflicts of Interest

The authors declare no financial or commercial conflicts of interest. The study team is based in the hospital where the research was conducted and has an academic and professional interest in the implementation of accelerometer-based monitoring in clinical care.

Multimedia Appendix 1
Cardiology ward description and implementation strategies.
[DOCX File, 17 KB - cardio_v10i1e92189_app1.docx]

Multimedia Appendix 2
Observed total daily physical activity (min) and model-fitted values from the linear mixed effects interrupted time series model across 2 implementation phases of a ward-based physical activity self-monitoring intervention in cardiology inpatients at the University Medical Centre Utrecht (UMC Utrecht), the Netherlands (phase 1: April 2024-September 2024; phase 2: September 2024-March 2025).
[PNG File, 81 KB - cardio_v10i1e92189_app2.png]

Multimedia Appendix 3
Covariate balance before and after inverse probability weighting.
[DOCX File, 15 KB - cardio_v10i1e92189_app3.docx]

References

1. Adsett JA, Mudge AM. Interventions to promote physical activity and reduce functional decline in medical inpatients: an umbrella review. J Am Med Dir Assoc 2024 Aug;25(8):105052. [doi: 10.1016/j.jamda.2024.105052] [Medline: 38830596]
2. Brown CJ, Redden DT, Flood KL, Allman RM. The underrecognized epidemic of low mobility during hospitalization of older adults. J Am Geriatr Soc 2009 Sep;57(9):1660-1665. [doi: 10.1111/j.1532-5415.2009.02393.x] [Medline: 19682121]
3. Drazich BF, Resnick B, Boltz M, et al. Factors associated with physical activity in hospitalized patients with dementia. J Aging Phys Act 2023 Aug 1;31(4):658-665. [doi: 10.1123/japa.2022-0210] [Medline: 36746152]
4. Fernandez N, Gore S, Benson S, Blackwood J. Use of AM-PAC “6 click” scores to predict discharge location post-hospitalization in adults with cardiovascular disease: a retrospective cohort study. Cardiopulm Phys Ther J 2020;31(4):152-158. [doi: 10.1097/CPT.0000000000000128]
5. Antunes EL, Costa BM, Sochodolak RC, Vargas LM, Okuno NM. The influence of physical activity level on the length of stay in hospital in older men survivors of COVID-19. Sport Sci Health 2022;18(4):1483-1490. [doi: 10.1007/s11332-022-00948-7] [Medline: 35730029]
6. Jawad BN, Petersen J, Andersen O, Pedersen MM. Variations in physical activity and sedentary behavior during and after hospitalization in acutely admitted older medical patients: a longitudinal study. BMC Geriatr 2022 Mar 15;22(1):209. [doi: 10.1186/s12877-022-02917-8] [Medline: 35291952]
7. Koenders N, Potkamp-Kloppers S, Geurts Y, Akkermans R, Nijhuis-van der Sanden MWG, Hoogeboom TJ. Ban bedcentricity: a multifaceted innovation to reduce sedentary behavior of patients during the hospital stay. Phys Ther 2021 Jul 1;101(7):pzab054. [doi: 10.1093/ptj/pzab054] [Medline: 33564890]

XSL•FO
RenderX

8. Browning L, Denehy L, Scholes RL. The quantity of early upright mobilisation performed following upper abdominal surgery is low: an observational study. Aust J Physiother 2007;53(1):47-52. [doi: 10.1016/s0004-9514(07)70061-2] [Medline: 17326738]
9. Bayoumy K, Gaber M, Elshafeey A, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug;18(8):581-599. [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
10. de Leeuwerk ME, Bor P, van der Ploeg HP, et al. The effectiveness of physical activity interventions using activity trackers during or after inpatient care: a systematic review and meta-analysis of randomized controlled trials. Int J Behav Nutr Phys Act 2022 May 23;19(1):59. [doi: 10.1186/s12966-022-01261-9] [Medline: 35606852]
11. Gerry S, Bonnici T, Birks J, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. BMJ 2020 May 20;369:m1501. [doi: 10.1136/bmj.m1501] [Medline: 32434791]
12. Hoffmann T, Bakhit M, Michaleff Z. Shared decision making and physical therapy: what, when, how, and why? Braz J Phys Ther 2022;26(1):100382. [doi: 10.1016/j.bjpt.2021.100382] [Medline: 35063699]
13. Mizrahi D, Lai JKL, Wareing H, et al. Effect of exercise interventions on hospital length of stay and admissions during cancer treatment: a systematic review and meta-analysis. Br J Sports Med 2024 Jan 3;58(2):97-109. [doi: 10.1136/bjsports-2023-107372] [Medline: 37989539]
14. Sourdet S, Lafont C, Rolland Y, Nourhashemi F, Andrieu S, Vellas B. Preventable iatrogenic disability in elderly patients during hospitalization. J Am Med Dir Assoc 2015 Aug 1;16(8):674-681. [doi: 10.1016/j.jamda.2015.03.011] [Medline: 25922117]
15. Szeto K, Arnold J, Singh B, Gower B, Simpson CEM, Maher C. Interventions using wearable activity trackers to improve patient physical activity and other outcomes in adults who are hospitalized: a systematic review and meta-analysis. JAMA Netw Open 2023 Jun 1;6(6):e2318478. [doi: 10.1001/jamanetworkopen.2023.18478] [Medline: 37318806]
16. van Grootel J, Bor P, Netjes JA, Veenhof C, Valkenet K. Improving physical activity in hospitalized patients: the preliminary effectiveness of a goal-directed movement intervention. Clin Rehabil 2023 Nov;37(11):1501-1509. [doi: 10.1177/02692155231189607] [Medline: 37487188]
17. Hoyer EH, Friedman M, Lavezza A, et al. A unit-based, multi-center evaluation of adopting mobility measures and daily mobility goals in the hospital setting. Appl Nurs Res 2023 Apr;70:151655. [doi: 10.1016/j.apnr.2022.151655] [Medline: 36933900]
18. Itoh H, Amiya E, Jimba T, et al. Efficacy and safety of remote cardiac rehabilitation in the recovery phase of cardiovascular diseases (RecRCR study): a multicenter, nonrandomized, and interventional trial in Japan. Int J Cardiol Heart Vasc 2024 Jun;52:101421. [doi: 10.1016/j.ijcha.2024.101421] [Medline: 38799401]
19. Nap HH, Stolwijk NE, Ipakchian Askari S, et al. The evaluation of a decision support system integrating assistive technology for people with dementia at home. Front Dement 2024;3:1400624. [doi: 10.3389/frdem.2024.1400624] [Medline: 39081617]
20. Iida H, Maehira H, Mori H, et al. Preoperative physical activity level measurement by accelerometer for predicting post-hepatectomy complications: a prospective observational study. Eur Surg Res 2023;64(2):193-200. [doi: 10.1159/000525280] [Medline: 35636396]
21. Iida H, Maehira H, Mori H, et al. Usefulness of measuring temporal changes in physical activity levels using an accelerometer for prediction and early detection of postoperative complications after hepatectomy. HPB (Oxford) 2022 Jan;24(1):57-64. [doi: 10.1016/j.hpb.2021.05.011] [Medline: 34158231]
22. Nevo Y, Shaltiel T, Constantini N, et al. Activity tracking after surgery: does it correlate with postoperative complications? Am Surg 2022 Feb;88(2):226-232. [doi: 10.1177/0003134820988818] [Medline: 33522277]
23. Kitamura M, Izawa KP, Nagasaki T, et al. Effects of self-monitoring using an accelerometer on physical activity of older people with long-term care insurance in Japan: a randomized controlled trial. Eur Geriatr Med 2024 Apr;15(2):371-380. [doi: 10.1007/s41999-024-00935-w] [Medline: 38353911]
24. Haley SM, Coster WJ, Andres PL, et al. Activity outcome measurement for postacute care. Med Care 2004 Jan;42(1 Suppl):I49-I61. [doi: 10.1097/01.mlr.0000103520.43902.6c] [Medline: 14707755]
25. Heinzmann J, Baumgartner C, Liechti FD. Goal-directed mobility of medical inpatients–a mini review of the literature. Front Med (Lausanne) 2022;9:878031. [doi: 10.3389/fmed.2022.878031] [Medline: 35665320]
26. Karakoc ZB, Kuru Colak T, Sari˙ Z, Polat MG. The effect of virtual rehabilitation added to an accelerated rehabilitation program after anterior cruciate ligament reconstruction: a randomized controlled trial. Clin Exp Health Sci 2019;9(2):124-129. [doi: 10.33808/clinexphealthsci.564273]
27. Shelton RC, Cooper BR, Stirman SW. The sustainability of evidence-based interventions and practices in public health and health care. Annu Rev Public Health 2018 Apr 1;39:55-76. [doi: 10.1146/annurev-publhealth-040617-014731] [Medline: 29328872]
28. Holtrop JS, Estabrooks PA, Gaglio B, et al. Understanding and applying the RE-AIM framework: clarifications and resources. J Clin Transl Sci 2021;5(1):e126. [doi: 10.1017/cts.2021.789] [Medline: 34367671]
29. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017 Feb 1;46(1):348-355. [doi: 10.1093/ije/dyw098] [Medline: 27283160]
30. Digitale JC, Stojanovski K, McCulloch CE, Handley MA. Study designs to assess real-world interventions to prevent COVID-19. Front Public Health 2021;9:657976. [doi: 10.3389/fpubh.2021.657976] [Medline: 34386470]

XSL•FO
RenderX

31. Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. BMJ Glob Health 2020 Oct;5(10):e003567. [doi: 10.1136/bmjgh-2020-003567] [Medline: 33055094]
32. Hassett L, Wolfenden L. Research note: designing implementation trials in physiotherapy. J Physiother 2022 Jul;68(3):210-212. [doi: 10.1016/j.jphys.2022.05.009] [Medline: 35725719]
33. Goodman D, Ogrinc G, Davies L, et al. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines, v2.0: examples of SQUIRE elements in the health care improvement literature. BMJ Qual Saf 2016 Dec;25(12):e7. [doi: 10.1136/bmjqs-2015-004480] [Medline: 27076505]
34. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
35. Valkenet K, Bor P, van Delft L, Veenhof C. Measuring physical activity levels in hospitalized patients: a comparison between behavioural mapping and data from an accelerometer. Clin Rehabil 2019 Jul;33(7):1233-1240. [doi: 10.1177/0269215519836454] [Medline: 30864490]
36. Adams C, Walpola R, Schembri AM, Harrison R. The ultimate question? Evaluating the use of Net Promoter Score in healthcare: a systematic review. Health Expect 2022 Oct;25(5):2328-2339. [doi: 10.1111/hex.13577] [Medline: 35985676]
37. Krol MW, de Boer D, Delnoij DM, Rademakers JJDM. The net promoter score—an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109. [doi: 10.1111/hex.12297] [Medline: 25345554]
38. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol 2018 Dec 1;47(6):2082-2093. [doi: 10.1093/ije/dyy135] [Medline: 29982445]
39. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, McKenzie JE. Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. BMC Med Res Methodol 2021 Jun 26;21(1):134. [doi: 10.1186/s12874-021-01306-w] [Medline: 34174809]
40. Compernolle S, DeSmet A, Poppe L, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63. [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
41. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008 Oct;13:64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
42. Geelen SJG, van Dijk-Huisman HC, de Bie RA, et al. Barriers and enablers to physical activity in patients during hospital stay: a scoping review. Syst Rev 2021 Nov 4;10(1):293. [doi: 10.1186/s13643-021-01843-x] [Medline: 34736531]
43. Thomas EM, Smith J, Curry A, et al. Association of physical function with hospital readmissions among older adults: a systematic review. J Hosp Med 2025 Mar;20(3):277-287. [doi: 10.1002/jhm.13538] [Medline: 39494712]
44. Ward-Stockham K, Omonaiye O, Darzins P, et al. Understanding the influences on hospital discharge decision-making from patient, carer and staff perspectives. BMC Health Serv Res 2024 Sep 19;24(1):1097. [doi: 10.1186/s12913-024-11581-0] [Medline: 39300431]
45. Gallardo-Gómez D, Del Pozo-Cruz J, Pedder H, et al. Optimal dose and type of physical activity to improve functional capacity and minimise adverse events in acutely hospitalised older adults: a systematic review with dose-response network meta-analysis of randomised controlled trials. Br J Sports Med 2023 Oct;57(19):1272-1278. [doi: 10.1136/bjsports-2022-106409] [Medline: 37536984]
46. Hoyer EH, Brotman DJ, Chan KS, Needham DM. Barriers to early mobility of hospitalized general medicine patients: survey development and results. Am J Phys Med Rehabil 2015 Apr;94(4):304-312. [doi: 10.1097/PHM.0000000000000185] [Medline: 25133615]
47. Valkenet K, McRae P, Reijneveld E, et al. Inpatient physical activity across a large university city hospital: a behavioral mapping study. Physiother Theory Pract 2024 Jan 2;40(1):153-160. [doi: 10.1080/09593985.2022.2112116] [Medline: 36036375]
48. Robinson TN, Carmichael H, Hosokawa P, et al. Decreases in daily ambulation forecast post-surgical re-admission. Am J Surg 2022 May;223(5):857-862. [doi: 10.1016/j.amjsurg.2021.08.011] [Medline: 34392912]
49. Grossglauser M, Saner H. Data-driven healthcare: from patterns to actions. Eur J Prev Cardiol 2014 Nov;21(2 Suppl):14-17. [doi: 10.1177/2047487314552755] [Medline: 25354949]
50. Sun J, Lin Q, Zhao P, et al. Reducing waiting time and raising outpatient satisfaction in a Chinese public tertiary general hospital-an interrupted time series study. BMC Public Health 2017 Aug 22;17(1):668. [doi: 10.1186/s12889-017-4667-z] [Medline: 28830400]
51. Thayer WM, Hasan MZ, Sankhla P, Gupta S. An interrupted time series analysis of the lockdown policies in India: a national-level analysis of COVID-19 incidence. Health Policy Plan 2021 Jun 3;36(5):620-629. [doi: 10.1093/heapol/czab027] [Medline: 33899097]
52. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393. [doi: 10.1136/bmj.b2393] [Medline: 19564179]

XSL•FO
RenderX

Abbreviations

BCT: behavior change techniques
EHR: electronic health record
GM: geometric mean
HCP: health care practitioner
IPW: inverse probability weighting
ITS: interrupted time series
MET: metabolic equivalent task
NPS: net promoter score
PAM: Physical Activity Monitor (PAM AM400)
RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance
SQUIRE: Standards for Quality Improvement Reporting Excellence
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
UMC: University Medical Centre
WMO: Dutch Medical Research Involving Human Subjects Act

Edited by A Coristine; submitted 27.Jan.2026; peer-reviewed by C Garcia-Conejo, K Kitagaki; revised version received 17.Aug.2026; accepted 19.Aug.2026; published 07.Oct.2026.

Please cite as:
de Miguel Llorente C, Sieswerda GTJ, Karramass K, Rommens C, Hoogeboom TJ, Valkenet K
Real-Time Physical Activity Self-Monitoring in Patients Hospitalized for Cardiology Care: Interrupted Time Series Analysis
JMIR Cardio 2026;10:e92189
URL: https://cardio.jmir.org/2026/1/e92189
doi:10.2196/92189

© Carlos de Miguel Llorente, Gertjan T J Sieswerda, Kaoutar Karramass, Charlotte Rommens, Thomas J Hoogeboom, Karin Valkenet. Originally published in JMIR Cardio (https://cardio.jmir.org), 7.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Use and Interest in Virtual Cardiac Rehabilitation Among Cardiac Inpatients: Cross-Sectional Study

Hamza G Patel[1,2], MD; Jie Ding[2], PhD; Xinyi Zhang[2,3], MSN, RN; Abel Cherian[2,4], MD; Pauline Hyunh[2,5], MD; Apurva Sharma[1,2], MD; Vennela Avula[1,2], MD; Rongzi Shan[2,6], MD; Katherine A Ornstein[3], PhD; Shireen R Khoury[1,2], MD, MPH; Francoise A Marvel[1,2], MD; Lena M Mathews[1,2], MD, MHS; Seth S Martin[1,2,7,8,9], MD, MHS; Erin M Spaulding[2,3,8], PhD, RN

[1]Division of Cardiology Department of Medicine, Johns Hopkins Medicine, Baltimore, MD, United States

[2]Digital Health Innovation Laboratory Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[3]School of Nursing, Johns Hopkins University, 525 North Wolfe Street, Baltimore, MD, United States

[4]University of Arizona College of Medicine-Tucson/Banner University Medical Center, Tucson, AZ, United States

[5]Kaiser Permanente Oakland Medical Center, Oakland, CA, United States

[6]Smidt Heart Institute Department of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[7]Center for Health Equity, Johns Hopkins University, Baltimore, MD, United States

[8]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, United States

[9]Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Erin M Spaulding, PhD, RN
Digital Health Innovation Laboratory Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Abstract

Background: Cardiac rehabilitation (CR) improves patient quality of life, morbidity, and mortality. Unfortunately, it is underused by patients. Digital health interventions offer a solution to increase participation in CR. However, patients' interest in virtual CR, especially among those in the inpatient setting, has not been fully explored. The benefits of CR have been predominantly demonstrated in traditional, center-based CR programs.

Objective: The objective of this cross-sectional study with prospective enrollment was to explore inpatient interest in virtual CR among adult patients who were hospitalized with a CR-qualifying diagnosis.

Methods: A Qualtrics survey comprising multiple-choice questions was administered to cardiac inpatients at the progressive cardiac care unit at the Johns Hopkins Hospital from January 2020 to March 2024. Sociodemographic and clinical characteristics were retrieved from the electronic medical records. The study included English-speaking patients aged 18 years or older with a diagnosis eligible for CR. Digital health use was compared between an earlier enrollment period (2020 - 2022) and a later enrollment period (2023 - 2024).

Results: A total of 150 patients were included (mean age 64, SD 13 years; n=57, 38% female, and n=85, 56.7% White). With respect to sociodemographic characteristics, 26% (n=39) of the patients had a high school education or less, 53.8% (70/130) were married, 26.2% (34/130) were employed full time, and 63.8% (83/130) had private insurance. In exploratory enrollment-period comparisons, texting health care providers was more commonly reported in 2023 to 2024 than in 2020 to 2022 (74/123, 60.2% vs 6/27, 22.2%; *P*<.001), as was using a smartphone or tablet to learn about illnesses (93/123, 75.6% vs 17/27, 63%; *P*=.04) and to discuss health with a physician or nurse practitioner (86/123, 69.9% vs 12/27, 44.4%; *P*=.01). Participants with more than a high school education were more likely to perceive smartphones as beneficial for leading a healthier lifestyle (50/104, 48.1% vs 9/37, 24.3%, *P*=.01) and learning about illnesses (90/105, 85.7% vs 20/37, 54.1%, *P*<.001) than were participants with a high school education or less. Participants across all sociodemographic factors expressed interest in virtual CR (overall 107/150, 71.3%), with non-White participants being more interested than White participants (55/65, 84.6% vs 52/85, 61.2%, *P*=.002).

Conclusions: Most cardiac inpatients expressed interest in home-based or virtual CR to alleviate barriers to in-person CR participation. Future work should emphasize digital equity and user support to optimize the widespread adoption of virtual CR.

(JMIR Cardio 2026;10:e91767) doi:10.2196/91767

XSL•FO
RenderX

KEYWORDS

digital health; telemedicine; virtual cardiac rehabilitation; cardiac rehabilitation; mobile health

Introduction

There are many effective interventions for cardiovascular disease (CVD) that resolve immediate risk and symptoms in patients. However, the modifiable risk factors that predispose patients to developing CVD in the first place remain to be fully addressed, and CVD remains the most significant cause of mortality worldwide [1]. Cardiac rehabilitation (CR) bridges the gap in the transition from acute interventions to long-term preventative care [2]. CR is a multidisciplinary intervention that includes health education and supervised physical activity tailored to the individual. This reduces the burden of CVD by enhancing the patient's recovery from an acute cardiac event and contributing to lasting lifestyle modifications. Many studies have demonstrated the benefits of CR, predominantly in traditional, center-based CR programs, including its ability to improve cardiovascular health and quality of life as well as decrease mortality and hospitalizations [3,4].

Despite this strong evidence, CR is widely underutilized by patients, with attendance rates in the United States ranging from 19% to 34% in national analyses [5,6]. Additionally, the 2019 United Kingdom National Audit of Cardiac Rehabilitation revealed that older individuals, women, patients with multiple comorbidities, patients in rural communities, and ethnic minority groups are much less likely to attend CR than others [7]. This lack of use is due to a variety of interrelated factors at the patient, clinician, and health care levels. At the patient level, the cost of traveling to the CR program, ease of accessibility, time away from work, and lack of motivation, among other reasons, tend to be common concerns that prevent patients from engaging in CR [4]. At the clinician level, a lack of introduction to CR in medical training could translate to an inability to properly emphasize the importance of CR during patient education [4]. Even when referral occurs, however, referral alone may be insufficient to ensure patient participation. Referral rates to CR programs by physicians increased to approximately 80% in more recent studies among patients after acute myocardial infarction, but less than half of the patients attended [5,8]. Finally, at the health care level, this lack of use of CR by patients could be due to reduced capacities, temporary closure of centers, and lack of reimbursement and resources [9,10].

These system-level barriers became particularly visible during the COVID-19 pandemic, when in-person, group-based CR sessions were disrupted by infection-control measures and shelter-in-place policies [9,10]. However, the pandemic also gave rise to the need for the extension of CR services to the home and to new avenues of delivering home-based CR to patients by leveraging technology, such as smartphones and wearable devices [9-12].

Much of the focus on home-based, virtual CR has been on rationalizing its utility as well as focusing on the physical activity and exercise component of CR [13]. A systematic review of randomized controlled trials comparing home-based (which may have included digital interventions) with center-based CR determined that home-based CR is noninferior to center-based CR in total mortality, exercise capacity, or health-related quality of life 24 months after CR [14]. Virtual CR has also been endorsed by the American Heart Association for its ability to overcome the accessibility barriers inherent to traditional, in-person, center-based CR [12]. However, how comfortable patients with CVD are with technology, what sociodemographic factors influence their technology use for health care purposes, and whether patients would be interested in using technology to facilitate CR at home have not been fully explored. A cross-sectional study conducted using the Health Information National Trends Survey found that 50% of US adults with CVD reported having a telehealth visit (video or audio - only) in the past year, with the majority being satisfied with their care [15]. Another cross-sectional study conducted in Israel, involving patients with CVD, found that 83% of the study population was interested in remote digital CR [16]. Younger age, sex, and education level did not impact this interest; however, it is unknown whether this would be the case with prolonged use of virtual CR [12,16].

This study aimed to assess the willingness of patients to engage in virtual CR. Specifically, we aimed to (1) describe patient awareness of, perceived benefits of, and barriers to attending traditional, in-person, center-based CR; (2) characterize technology ownership and digital health use across earlier and later enrollment periods, broadly corresponding to pandemic-era and postacute pandemic care patterns; and (3) evaluate patient interest in virtual CR and identify sociodemographic and clinical factors associated with interest.

Methods

Study Design and Participants

In this cross-sectional study with prospective enrollment, we enrolled patients admitted to the progressive cardiac care unit (PCCU) at Johns Hopkins Hospital. Johns Hopkins has an in-person, center-based CR program, but at the time of recruitment for this study, it did not offer virtual CR as part of usual care. Potentially eligible participants were identified based on the following CR-qualifying diagnoses and interventions: acute myocardial infarction (ST-elevation myocardial infarction [STEMI] or non-STEMI [NSTEMI] type 1), angina, heart failure, coronary artery bypass grafting (CABG), coronary artery angioplasty or percutaneous coronary intervention (PCI), heart valve surgery, transcatheter aortic valve replacement (TAVR), or heart transplantation [17]. The inpatient PCCU was selected as the survey setting because hospitalization for a CR-qualifying diagnosis represents a key predischarge opportunity for CR education, referral, and assessment of patient preferences regarding center-based, home-based, and virtual rehabilitation options. Patients were excluded if they declined participation or if the primary team felt they were not appropriate for bedside survey administration because of clinical instability, inability to provide informed consent, cognitive impairment, or another clinical factor that would preclude participation. These decisions

XSL•FO
RenderX

were made based on the clinical judgment of the treatment team rather than by a formalized exclusion checklist. Participants were recruited from January 2020 to March 2024. Recruitment was not formally paused during the pandemic; however, in-person recruitment was limited during portions of the earlier study period because of pandemic-related restrictions on bedside research activities and patient contact.

For exploratory temporal analyses of digital health use, participants were grouped into an earlier enrollment period (2020 - 2022) and a later enrollment period (2023 - 2024). This cutoff was chosen to broadly distinguish patients enrolled during the earlier pandemic-era implementation of telehealth and digital health services from those enrolled after the acute pandemic period and near the end of the COVID-19 public health emergency in 2023 [18]. Because enrollment was imbalanced between these periods, temporal comparisons were considered exploratory. A complete screening log of all CR-eligible patients admitted to the PCCU, patients approached for participation, and patients who declined participation was not maintained throughout the full enrollment period. Therefore, the total number of eligible patients during the study period, the number approached, and the proportion who declined participation could not be reliably determined. Study team members approached patients at the bedside in the PCCU to confirm their eligibility, determine their interest in the study, obtain informed consent, and administer the survey on a tablet or laptop via Qualtrics.

Ethical Considerations

This study was approved by the Johns Hopkins University School of Medicine Institutional Review Board (IRB00221052), and a waiver of documentation of consent was obtained.

Measures and Data Collection

A survey developed by the study team and based on the framework of the American Medical Association Digital Health Survey (Multimedia Appendix 1) was administered to the participants in person [19]. The survey comprised multiple-response questions (ie, select all that apply) to collect data on participant demographics (age, sex, race, and educational attainment); technology ownership and past digital health use; knowledge of, attitudes toward, and barriers to attending CR; and attitudes toward virtual CR. In the survey, participants were provided with an explanation of what CR entails, including the length of the program and the various components. The survey was designed for this study and was not formally psychometrically validated. Data were also extracted from participants' electronic medical records regarding their marital status, employment status, health insurance status, CR-eligible diagnoses and/or interventions, past attendance at CR, and number of hospitalizations in the past 5 years.

Statistical Analysis

Descriptive statistics were summarized as means with SDs for continuous data with a normal distribution, medians with IQRs for continuous data with a skewed distribution, and frequencies with percentages for categorical data. Chi-square or Fisher exact tests were used to compare technology ownership and digital health use among participants enrolled before or during (2020 - 2022) and after (2023 - 2024) the COVID-19 pandemic. Given the imbalance in group sample sizes, these temporal comparisons were prespecified as exploratory and were not intended to establish definitive changes attributable to the COVID-19 pandemic [18]. Sociodemographic characteristics of participants (ie, age, sex, race, educational attainment, marital status, employment status, and health insurance status) were presented according to digital health use status, including technology ownership, digital health use, and perceived smartphone benefits to health. Both sociodemographic and clinical characteristics (number of prior hospitalizations in the past 5 years, prior CR attendance, number of qualifying CR diagnoses, and type of qualifying CR diagnosis [medical only, procedural only, or both]) of participants were presented according to CR-related responses by participants including awareness of CR, perceived benefits of CR, barriers to attending CR, and interest in virtual CR. Medical CR-qualifying diagnoses included acute myocardial infarction and/or heart failure, and procedural qualifying CR diagnoses included CABG, PCI, TAVR, and/or heart transplant. Those sociodemographic and clinical characteristics were compared between dichotomized groups of digital health use status and CR-related responses using independent 2-tailed *t* tests or Mann-Whitney *U* tests for continuous data and chi-square or Fisher exact tests for categorical variables.

Complete-case analysis was used for the primary analyses, with denominators reported based on available data for each outcome or covariate. Missing data were primarily due to item-level survey nonresponse and missing electronic medical record–derived variables among early-study participants for whom identifying information was not collected. Because missingness was related in part to early-study data collection procedures, the assumption that data were missing completely at random may not fully hold. Multiple imputation was not performed because the primary outcome, interest in virtual CR, was available for all participants, and missingness primarily affected secondary sociodemographic and clinical subgroup variables. Therefore, subgroup analyses using variables with missing data were interpreted as exploratory.

No formal a priori sample size calculation was performed. The analytic sample represented a consecutive convenience sample of eligible cardiac inpatients who were identified by electronic medical record review, approached by the study team when feasible, and agreed to participate during the enrollment period. No adjustment for multiple comparisons was applied because these subgroup analyses were exploratory and intended to generate hypotheses rather than test a prespecified hierarchy of confirmatory hypotheses. Given the modest sample size and limited number of events in some subgroups, multivariable modeling was not performed for all outcomes because of concern about model overfitting. Therefore, subgroup findings should be interpreted as exploratory and hypothesis generating.

Results

Overview

A total of 150 patients were included (mean age 64, SD 13 years; n=57, 38% female; and n=85, 56.7% White). With respect to sociodemographic characteristics, 26% (39/150) of the

XSL•FO
RenderX

patients had a high school education or less, 46.7% (70/150) were married, 26.2% (34/130) were employed full time, and 63.8% (83/130) had private insurance (Table 1).

XSL•FO
RenderX

Table . Baseline characteristics (N=150).

Characteristics	Participants
Sociodemographic characteristics	
Age (y), mean (SD)	64.1 (13.0)
Sex, n (%)	
Male	93 (62)
Female	57 (38)
Race, n (%)	
Asian	6 (4)
Black or African American	51 (34)
White	85 (56.7)
Other	8 (5.3)
Education, n (%)	
High school or less	39 (26)
Greater than high school	111 (74)
Marital status (n=130), n (%)[a]	
Married	70 (53.8)
Unmarried or other	60 (46.2)
Employment status (n=130), n (%)[a]	
Full time or self-employed	34 (26.2)
Other	96 (73.8)
Insurance status (n=130), n (%)[a]	
Private or other	83 (63.8)
Medicare or Medicaid	47 (36.2)
Clinical characteristics	
Number of prior hospitalizations in past 5 years, median (IQR)[a]	3 (2-5)
Prior CR[b] participation (n=130), n (%)[a]	42 (32.3)
Index hospitalization length of stay (d), median (IQR)[a]	9 (4-16)
Admitted to critical care during index hospitalization (n=130), n (%)[a]	34 (26.2)
Number of CR-qualifying diagnoses or interventions, median (IQR)[a,c]	2 (1-2)
CR-eligible diagnoses or interventions (n=130), n (%)[a,c]	
ST-elevation myocardial infarction	23 (17.7)
Non–ST-elevation myocardial infarction type 1	19 (14.6)
Angina	21 (16.2)
Heart failure	85 (65.4)
Coronary artery bypass grafting	26 (20)
Percutaneous coronary intervention	41 (31.5)
Heart valve surgery	20 (15.4)
Transcatheter aortic valve replacement	18 (13.8)
Heart transplant	10 (7.7)
Type of qualifying diagnosis or intervention (n=130), n (%)[a]	
Medical diagnosis only	42 (32.3)

XSL•FO
RenderX

Characteristics	Participants
Procedural diagnosis or intervention only	23 (17.7)
Both medical and procedural qualifying diagnoses or interventions	65 (50)

[a]Denominators vary because of item-level survey nonresponse and missing electronic medical record–derived variables for early-study participants for whom identifying information was not collected. Percentages were calculated among participants with available data for each variable.

[b]CR: cardiac rehabilitation.

[c]Participants may have had more than one qualifying CR diagnosis or intervention; therefore, the percentages exceed 100%.

Interest in Virtual CR

Overall, of 150 participants, 107 (71.3%) expressed interest in virtual CR. Interest in virtual CR was higher among non-White participants than among White participants (55/65, 84.6% vs 52/85, 61.2%; *P*=.002). Interest in virtual CR did not differ significantly by age, sex, education, marital status, employment status, health insurance status, number of prior hospitalizations, number of qualifying diagnoses, type of qualifying diagnosis, or prior CR participation (Table 2). Because multiple subgroup comparisons were performed, these findings should be interpreted as exploratory.

XSL•FO
RenderX

Table . Sociodemographic and clinical associations with interest in virtual cardiac rehabilitation (CR)[a].

	No interest in virtual CR	Interest in virtual CR	*P* value[b]
Age (y), mean (SD)	65.7 (13.9)	63.5 (12.7)	.35
Sex, n (%)			.38
Female (n=57)	14 (24.6)	43 (75.4)	
Male (n=93)	29 (31.2)	64 (68.8)	
Race, n (%)			.002
White (n=85)	33 (38.8)	52 (61.2)	
Non-White (n=65)	10 (15.4)	55 (84.6)	
Education, n (%)			.94
High school or less (n=39)	11 (28.2)	28 (71.8)	
Greater than high school (n=111)	32 (28.8)	79 (71.2)	
Marital status, n (%)			.86
Married (n=70)	20 (28.6)	50 (71.4)	
Unmarried (n=60)	18 (30)	42 (70)	
Employment status, n (%)			.37
Full time or self-employed (n=34)	35.3 (12)	64.7 (22)	
Other (n=96)	27.1 (26)	72.9 (70)	
Health insurance, n (%)			.49
Private or other (n=83)	26 (31.3)	57 (68.7)	
Medicare or Medicaid (n=47)	12 (25.5)	35 (74.5)	
Number of prior hospitalizations, median (IQR)	3 (2-5)	3 (2-5)	.89
Sum of qualifying diagnoses, median (IQR)	2 (1-3)	2 (1-2)	.43
Type of qualifying diagnosis, n (%)			.06
Medical diagnosis (n=42)	35 (83.3)	7 (16.7)	
Procedural diagnosis (n=23)	13 (56.5)	10 (43.5)	
Both diagnoses (n=65)	44 (67.7)	21 (32.3)	
Prior CR, n (%)			.26
No prior rehabilitation (n=88)	65 (73.9)	23 (26.1)	
Prior CR (n=42)	27 (64.3)	15 (35.7)	

[a]Denominators vary because of item-level survey nonresponse and missing electronic medical record–derived variables for early-study participants for whom identifying information was not collected. Percentages were calculated among participants with available data for each variable.

[b]*P* values were calculated using chi-square or Fisher exact tests for categorical variables, independent *t* tests for normally distributed continuous variables, and Mann-Whitney *U* tests for nonnormally distributed continuous variables, as appropriate.

Technology Ownership and Digital Health Use

Among the 150 participants, 27 (18%) were enrolled during the earlier enrollment period from 2020 to 2022 and 123 (82%) were enrolled during the later enrollment period from 2023 to 2024. Given this imbalance, comparisons between enrollment periods should be interpreted as exploratory. Smartphone ownership was high in both periods, with 100% (n=27) of participants enrolled in 2020 to 2022 owning a smartphone and 94.3% (115/122) of those enrolled in 2023 to 2024 reporting ownership. The use of smartwatches and wearables was higher in the 2023 to 2024 cohort (40/122, 32.8%) than in the 2020 to 2022 cohort (7/27, 25.9%), although the differences were not statistically significant (*P*=.48).

Some forms of digital health use were higher after the COVID-19 pandemic. Notably, the proportion of participants who texted their health care providers was significantly higher after the pandemic (2020 - 2022: 6/27, 22.2%; 2023 - 2024: 74/123, 60.7%; *P*<.001). Other forms of digital health use, including health app use (2020 - 2022: 11/27, 40.7%; 2023 - 2024: 57/123, 49.6%; *P*=.68), were similar, while sharing information from health apps with health care providers (2020 - 2022: 25.9%; 2023 - 2024: 45.5%; *P*=.08) showed a

XSL•FO
RenderX

numerical but nonsignificant increase after the pandemic. Posting health information on social media platforms (2020 - 2022: 3/27, 11.1%; 2023 - 2024: 6/123, 4.9%; *P*=.37) and watching health-related videos on YouTube (Google Inc; 2020 - 2022: 12/27, 44.4%; 2023 - 2024: 60/123, 48.8%; *P*=.17) were not significantly different after the pandemic (Table 3).

Table . Digital health use in participants enrolled before, during, and after the COVID-19 pandemic.

	Enrolled before or during the COVID-19 pandemic (2020 - 2022; n=27), n (%)	Enrolled after the COVID-19 pandemic (2023 - 2024; n=123), n (%)	*P* value[a]
Health app use	11 (40.7)	57 (46.3)	.68
Shared information from health apps with HCP[b]	7 (25.9)	55 (44.7)	.08
Texted HCP	6 (22.2)	74 (60.2)	<.001
Posted health information on social media	3 (11.1)	6 (4.9)	.37
Watched a health-related video on YouTube	12 (44.4)	60 (48.8)	.17

[a]*P* values were calculated using chi-square or Fisher exact tests.

[b]HCP: health care provider.

The proportion of participants who reported that their smartphone assisted in leading a healthier lifestyle was numerically higher after the pandemic (2020 - 2022: 29.6%; 2023 - 2024: 41.5%; *P*=.15), though this difference did not reach statistical significance. The use of smartphones for learning about illnesses (2020 - 2022: 62.3%; 2023 - 2024: 75.6%; *P*=.04) and discussing health with health care providers (2020 - 2022: 44.4%; 2023 - 2024: 69.9%; *P*=.01) were significantly higher after the pandemic (Table 4).

Table . Perceived smartphone benefits to health in participants enrolled before, during, and after the COVID-19 pandemic (N=150).

	Enrolled before or during the COVID-19 pandemic (2020-2022; n=27), n (%)	Enrolled after the COVID-19 pandemic (2023-2024; n=123), n (%)	*P* value[a]
Smartphone or tablet assisted in living a healthier lifestyle	8 (29.6)	51 (41.5)	.15
Smartphone or tablet assisted in learning about illnesses	17 (63)	93 (75.6)	.04
Smartphone or tablet assisted in talking about health with provider	12 (44.4)	86 (69.9)	.01

[a]*P* values were calculated using chi-square or Fisher exact tests.

Predictors of Technology Ownership and Digital Health Use

Smartphone and tablet ownership was consistent across age, sex, race, marital status, education, and employment status. Smartwatch and wearable ownership was also consistent across age, sex, race, marital status, and employment status. However, participants with greater than a high school education had higher rates of smartwatch or wearable device ownership than those with only a high school education or less (41/110, 37.3% vs 6/39, 15.4%; *P*=.01; Multimedia Appendix 2).

White participants were more likely to engage in texting health care providers compared with non-White participants (55/85, 64.7% vs 25/64, 39.1%; *P*=.002). Similarly, participants with greater than a high school education were more likely to use health apps (68/114, 59.4% vs 9/37, 24.3%; *P*=.001) and watch YouTube health videos (59/111, 53.2% vs 13/39, 33.3%; *P*=.03) than participants with a high school education or less. Texting health care providers was numerically higher among participants with greater than a high school education, but this difference was not statistically significant (*P*=.07). Married participants were more likely to use a health app (60.6% vs 33.9%; *P*=.003), share information from health apps with their health care providers (37/70, 52.9% vs 20/58, 34.5%; *P*=.04), and text their health care providers (50/70, 71.4% vs 27/59, 45.8%; *P*=.003) than their unmarried counterparts (Multimedia Appendix 2).

Participants with greater than a high school education were more likely to perceive smartphones as beneficial for leading a healthier lifestyle (50/104, 48.1% vs 9/37, 24.3%; *P*=.01) and learning about illnesses (90/105, 85.7% vs 20/37, 54.1%; *P*<.001) than participants with a high school education or less. Similarly, married participants were more likely to report that smartphones assisted with learning about illnesses (58/66, 87.9% vs 40/56, 71.4%; *P*=.02) and discussing their health with their health care providers (57/66, 86.4% vs 32/55, 58.2%; *P*<.001) than their unmarried counterparts. White participants were more likely to report that smartphones assisted with discussing their health with their health care providers (62/81, 76.5% vs 36/60, 60%; *P*=.04). Individuals employed full time (30/31, 96.8% vs 68/91, 74.7%; *P*=.008) were more likely to report that

XSL•FO
RenderX

smartphones assisted with learning about illnesses than their counterparts (Multimedia Appendix 2).

Awareness of and Barriers to Attending In-Person CR

Awareness of CR (106/149, 71.1%) did not differ by age, sex, race, education, marital status, employment status, number of prior hospitalizations, or number or type of CR-qualifying diagnoses. However, individuals with Medicare or Medicaid insurance were more likely to be aware of CR than individuals with private or other types of insurance (41/47, 87.2% vs 48/82, 58.5%; *P*=.001; Multimedia Appendix 2).

Non-White participants were more likely to report that CR could improve their sense of well-being than White participants (51/57, 89.5% vs 46/62, 74.2%; *P*=.03). Participants not employed full time (74/77, 96.1% vs 21/25, 84%; *P*=.04) were more likely to perceive CR as beneficial for preventing their condition from worsening than their counterparts. Perceived benefit of CR for preventing worsening of disease did not differ significantly by education level. Patients with no prior CR experience were more likely to believe that participating in CR would lower their chances of another cardiac event (69/71, 97.2% vs 26/31, 83.9%; *P*=.01). Individuals with a medical CR-qualifying diagnosis only were more likely to perceive CR as beneficial for preventing their condition from worsening (33/34, 97.1% vs 11/14, 78.6% with a procedural CR-qualifying diagnosis only vs 51/54, 94.4% with both; *P*=.01; Multimedia Appendix 2).

Younger patients (*P*=.001) and males (23/84, 27.4% vs 6/50, 12%; *P*=.04) were more likely to report absence from work as a barrier to attending CR. Full-time employees were more likely to report absence from work (17/33, 51.5% vs 10/87, 11.5%; *P*<.001) and lack of time (16/33, 48.5% vs 24/87, 27.6%; *P*=.03) as barriers to attending CR than their counterparts who were not employed full-time. Full-time employees were less likely to report travel (12/33, 36.4% vs 54/87, 62.1%; *P*=.01) and paperwork (3/33, 9.1% vs 10/87, 11.5%; *P*=.01) as barriers to attending CR than their counterparts who were not employed full-time. Although the difference in paperwork as a barrier reached statistical significance, the absolute difference was small and is unlikely to be clinically meaningful. Patients with a higher number of prior hospitalizations cited travel as a barrier to attending CR (*P*=.02). Individuals with a medical CR-qualifying diagnosis only were more likely to report cost as a barrier to attending CR (19/40, 47.5% vs 5/19, 26.3% with procedural CR-qualifying diagnosis only vs 15/61, 24.6% with both; *P*=.046; Multimedia Appendix 2).

Discussion

Principal Results

This cross-sectional study with prospective enrollment explored patient interest in virtual CR by assessing patients' attitudes toward and barriers to in-person CR, technology ownership, and digital health use. The results showed that while patients understood the benefits of CR, many cited travel, time, absence from work, and paperwork as barriers to attending CR. Barriers related to absence from work and time constraints were noted more frequently among younger patients, men, and full-time employees. Notably, most participants (107/150, 71.3%) expressed interest in virtual CR, especially non-White participants and those with procedural CR-qualifying diagnoses only. The use of digital health tools, including texting health care providers and sharing health data via apps, increased after the COVID-19 pandemic, suggesting a growing readiness among patients to adopt virtual health care solutions.

Comparison With Prior Work

Our findings are consistent with previous research showing that patients are interested in using virtual CR, especially to address barriers related to travel and time constraints [20]. Multiple studies have already found that home-based CR is as beneficial as center-based CR [14,21,22]. For clarity, center-based CR refers to traditional facility-based programs [10]. Home-based CR involves recovery at home with periodic supervision, and virtual CR is conducted through digital technologies such as telehealth or mobile apps [10]. Additionally, multiple studies have demonstrated that virtual or digital CR may safely facilitate home-based CR and improve adherence by reducing barriers to attendance, such as transportation costs, travel time to CR centers, and capacity constraints of centers [8,13,23,24]. The American Association of Cardiovascular and Pulmonary Rehabilitation, American Heart Association, and American College of Cardiology have stated that home-based CR can reasonably take the place of center-based CR for low- to moderate-risk patients, and digital health interventions can assist in cementing this transition [25].

A prior analysis among persons with atherosclerotic CVD found that health information technology (HIT) use was higher among women, younger adults, White adults, and individuals with a higher income, who were employed, married, and held a bachelor's degree or higher [26]. HIT use was defined differently in this analysis as the combination of looking up health information on the internet, filling a web-based prescription, scheduling a medical appointment on the internet, communicating with health care providers through email, or using web-based group chats to learn about health topics [26]. Despite differences between the prior NHIS analysis and the present study in how HIT or digital health use was defined, we similarly found that White participants, participants with higher educational attainment, and married participants used digital health tools more frequently.

Notably, non-White participants in our study were more interested in a virtual CR than White participants. Prior studies have described narrowing gaps in digital health use among historically underrepresented racial and ethnic groups, particularly in general telehealth and patient portal use, rather than in virtual CR specifically [27,28]. Our finding extends this broader digital health trend to the CR setting and suggests that virtual CR may be acceptable to patient groups that have historically faced barriers to traditional center-based CR. Further research is needed to determine whether this interest translates into virtual CR enrollment, adherence, and completion and to identify implementation strategies that support equitable access.

Nevertheless, these findings highlight the importance of ensuring that virtual CR platforms are designed to be accessible to all patients, regardless of race, age, gender, income level, or digital literacy. Barriers to CR are an important consideration for its

XSL•FO
RenderX

consistent use in patient populations. The usability and reliability of virtual CR interfaces, the support systems patients have access to when needing help with technology, the consideration of frailty in older adult patients in a nonclinical setting, and the lasting psychosocial effects of undergoing an adverse cardiac event are all potential challenges that must be addressed before virtual CR can be fully accepted by patients and physicians [12,29-32].

Our findings suggest that many cardiac inpatients are receptive to virtual CR, particularly in the context of commonly reported barriers to traditional, in-person, center-based CR. Rather than replacing center-based CR, virtual and hybrid CR models may serve as complementary approaches that expand access for patients who face barriers related to travel, time, work obligations, or other logistical constraints. These findings support further evaluation of virtual CR implementation, usability, adherence, and clinical outcomes in larger, multicenter studies. Importantly, efforts to expand virtual CR should also address digital equity, including access to devices, internet connectivity, digital literacy, language accessibility, and technical support.

Strengths and Limitations

This study has several strengths, including its timing before, during, and after the COVID-19 pandemic, which allowed us to capture patient digital health use during different periods. Additionally, our focus on a diverse population with a range of CR-qualifying cardiac conditions provides valuable insights into how different patient subgroups may approach virtual CR.

Limitations include the relatively small sample size and the conduct of the study at a single academic medical center, limiting our ability to conduct multivariate modeling as well as the generalizability of the findings. In addition, the manuscript includes multiple subgroup comparisons without statistical correction for multiple testing, which increases the risk of type I error. These analyses were exploratory and should be interpreted as hypothesis generating rather than confirmatory.

The enrollment-period comparison was also limited by substantial imbalance in group size, with 27 participants enrolled in 2020 - 2022 and 123 participants enrolled in 2023 - 2024. This imbalance limited statistical power for temporal comparisons and increased the risk that observed differences reflect recruitment patterns, pandemic-related disruptions, or secular trends in digital health adoption rather than true differences attributable to the COVID-19 pandemic. Therefore, these comparisons should be interpreted as exploratory and not as definitive evidence of pandemic-related changes.

In addition, because a complete screening log was not maintained, we could not determine how many eligible patients were admitted during the study period, how many were approached, or how many declined participation. This limits our ability to assess the recruitment rate and nonresponse bias and raises the possibility of selection bias, particularly if patients who were more clinically stable, more comfortable with technology, or more interested in CR were more likely to participate.

The survey instrument was developed for this study and was not formally psychometrically validated. Therefore, responses may be subject to measurement error, and self-reported interest in virtual CR may not fully predict actual postdischarge enrollment, adherence, or completion. Because surveys were administered at the bedside by study team members, responses may have been subject to social desirability bias, potentially inflating reported interest in virtual CR or perceived benefits of CR. Future studies should consider anonymous or self-administered survey formats to reduce this potential bias. Digital health attitudes during the study period may also have been influenced by pandemic-driven adaptation to telehealth and remote care rather than reflecting stable long-term preferences. Because this study assessed attitudes at a single time point during hospitalization, reported interest in virtual CR may not translate into actual postdischarge engagement, enrollment, adherence, or completion. Future longitudinal studies should assess whether inpatient interest in virtual CR persists after discharge and should evaluate CR enrollment, participation, and completion at clinically relevant follow-up points, including 6 and 12 months.

Another limitation is that identifying information was not collected for respondents enrolled at the beginning of the study; consequently, we were unable to collect complete sociodemographic and clinical characteristics from the electronic health record for these individuals, resulting in approximately 13% missing data for certain variables. Additionally, since the population surveyed comprised patients on the PCCU, the sample may not be representative of all cardiac patients admitted to general medicine floors or other locations. Additionally, participants were enrolled at one academic medical center with strong CR infrastructure, which may not reflect practices at other institutions. Because enrollment was limited to English-speaking patients, our findings may not be transferable to patients with limited English proficiency or to linguistically diverse populations. This limitation is particularly relevant to the equity implications of virtual CR, as language access may influence digital health use, patient education, and engagement with rehabilitation programs. Moreover, race was analyzed as White versus non-White, which aggregates heterogeneous racial and ethnic groups and may obscure important differences in digital health use, barriers to CR, and interest in virtual CR across specific populations. Future studies with larger and more diverse samples should disaggregate racial and ethnic categories to better understand subgroup-specific barriers, preferences, and implementation needs.

Finally, while we captured a range of sociodemographic and clinical characteristics, other factors such as health literacy, digital literacy, quality of internet access, data plan limitations, comfort with troubleshooting technology, and access to caregiver or technical support were not captured. Therefore, smartphone ownership should not be interpreted as equivalent to digital readiness, and our findings may overestimate patients' ability to successfully engage with virtual CR without additional support.

XSL•FO
RenderX

Conclusions

Overall, our findings suggest that patients with CR-eligible diagnoses face substantial barriers to attending in-person CR but are interested in using virtual CR to overcome logistical barriers. Given the single-site design and modest sample size, further research is needed to confirm the generalizability of these findings across care settings. Investigating why non-White patients showed greater interest in virtual CR is an important direction for future research.

Funding

The present study received no specific funding. Outside of this work, the authors received research support from the American Heart Association Health Technologies and Innovation Strategically Focused Research Network, a collaborative project of this network, and additional American Heart Association support. The authors also received funding support from the Patient-Centered Outcomes Research Institute, the National Institutes of Health (NIH), the David and June Trone Family Foundation, the Pollin Digital Innovation Fund, Sandra and Larry Small, Google, and Merck.

Data Availability

The deidentified data sets generated and analyzed during this study are not publicly available because they contain patient-level survey and electronic medical record data, and public sharing is limited by institutional review board and patient privacy considerations. Deidentified data may be available from the corresponding author on reasonable request and with appropriate institutional approvals.

Conflicts of Interest

Under a license agreement between Corrie Health and the Johns Hopkins University, the university owns equity in Corrie Health. SSM and FAM are entitled to royalty distributions related to technology. Additionally, SSM and FAM are cofounders of and hold equity in Corrie Health. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. SSM has received material support from Apple, iHealth, and Google. SSM is on the advisory board for Care Access and reports personal consulting fees from Amgen, Arrowhead, AstraZeneca, BMS, Chroma, Heartflow, Kaneka, NewAmsterdam, Novartis, Novo Nordisk, Premier, Sanofi, and Verve Therapeutics. FAM have received material support from Apple and iHealth. ES serves as a consultant to Corrie Health. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Patient survey.
[DOCX File, 39 KB - cardio_v10i1e91767_app1.docx]

Multimedia Appendix 2
Supplemental exploratory subgroup analyses of technology ownership, digital health use, perceived smartphone benefits, cardiac rehabilitation awareness, perceived cardiac rehabilitation benefits, and barriers to attending cardiac rehabilitation.
[DOCX File, 73 KB - cardio_v10i1e91767_app2.docx]

References

1. Global Burden of Cardiovascular Diseases and Risks 2023 Collaborators. Global, regional, and national burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990-2023. J Am Coll Cardiol 2025 Dec 2;86(22):2167-2243. [doi: 10.1016/j.jacc.2025.08.015] [Medline: 40990886]
2. Lee KC, Breznen B, Ukhova A, Koehler F, Martin SS. Virtual healthcare solutions for cardiac rehabilitation: a literature review. Eur Heart J Digit Health 2023 Feb;4(2):99-111. [doi: 10.1093/ehjdh/ztad005] [Medline: 36974268]
3. Dibben GO, Faulkner J, Oldridge N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 7;44(6):452-469. [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
4. Taylor RS, Dalal HM, McDonagh ST. The role of cardiac rehabilitation in improving cardiovascular outcomes. Nat Rev Cardiol 2022 Mar;19(3):180-194. [doi: 10.1038/s41569-021-00611-7] [Medline: 34531576]
5. Ritchey MD, Maresh S, McNeely J, et al. Tracking cardiac rehabilitation participation and completion among Medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020 Jan;13(1):e005902. [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
6. Martin SS, Aday AW, Allen NB, et al. 2025 Heart Disease and Stroke Statistics: a report of US and global data from the American Heart Association. Circulation 2025 Feb 25;151(8):e41-e660. [doi: 10.1161/CIR.0000000000001303] [Medline: 39866113]
7. National Audit of Cardiac Rehabilitation (NACR) quality and outcomes report 2019. : British Heart Foundation; 2019 URL: https://www.bhf.org.uk/informationsupport/publications/statistics/national-audit-of-cardiac-rehabilitation-quality-and-outcomes-report-2019 [accessed 2026-07-14]

XSL•FO
RenderX

8. Thomas RJ, Balady G, Banka G, et al. 2018 ACC/AHA clinical performance and quality measures for cardiac rehabilitation: a report of the American College of Cardiology/American Heart Association Task Force on performance measures. Circ Cardiovasc Qual Outcomes 2018 Apr;11(4):e000037. [doi: 10.1161/HCQ.0000000000000037] [Medline: 29599285]
9. Babu AS, Arena R, Ozemek C, Lavie CJ. COVID-19: a time for alternate models in cardiac rehabilitation to take centre stage. Can J Cardiol 2020 Jun;36(6):792-794. [doi: 10.1016/j.cjca.2020.04.023] [Medline: 32344000]
10. Thomas RJ. Cardiac rehabilitation - challenges, advances, and the road ahead. N Engl J Med 2024 Feb 29;390(9):830-841. [doi: 10.1056/NEJMra2302291] [Medline: 38416431]
11. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]
12. Golbus JR, Lopez-Jimenez F, Barac A, et al. Digital technologies in cardiac rehabilitation: a science advisory from the American Heart Association. Circulation 2023 Jul 4;148(1):95-107. [doi: 10.1161/CIR.0000000000001150] [Medline: 37272365]
13. Wongvibulsin S, Habeos EE, Huynh PP, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021 Feb 8;23(2):e18773. [doi: 10.2196/18773] [Medline: 33555259]
14. McDonagh ST, Dalal H, Moore S, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2023 Oct 27;10(10):CD007130. [doi: 10.1002/14651858.CD007130.pub5] [Medline: 37888805]
15. Bhatla A, Ding J, Mhaimeed O, et al. Patterns of telehealth visits after the COVID-19 pandemic among individuals with or at risk for cardiovascular disease in the United States. J Am Heart Assoc 2024 Sep 3;13(17):e036475. [doi: 10.1161/JAHA.124.036475] [Medline: 39206726]
16. Nabutovsky I, Nachshon A, Klempfner R, Shapiro Y, Tesler R. Digital cardiac rehabilitation programs: the future of patient-centered medicine. Telemed J E Health 2020 Jan;26(1):34-41. [doi: 10.1089/tmj.2018.0302] [Medline: 30789311]
17. Cardiac rehabilitation: an individualized supervised program for you. American Association of Cardiovascular and Pulmonary Rehabilitation (AACVPR). URL: https://www.aacvpr.org/Portals/0/Docs/AACVPR_CR_Fact_Sheet.pdf [accessed 2026-07-14]
18. CDC Museum COVID-19 timeline. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/museum/timeline/covid19.html [accessed 2026-07-14]
19. AMA digital health research. American Medical Association. 2020. URL: https://www.ama-assn.org/system/files/2020-02/ama-digital-health-study.pdf [accessed 2026-07-14]
20. Lear SA. The delivery of cardiac rehabilitation using communications technologies: the “virtual” cardiac rehabilitation program. Can J Cardiol 2018 Oct;34(10 Suppl 2):S278-S283. [doi: 10.1016/j.cjca.2018.07.009] [Medline: 30274638]
21. Anderson L, Sharp GA, Norton RJ, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6(6):CD007130. [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
22. Nso N, Nassar M, Mbome Y, et al. Comparative assessment of the long-term efficacy of home-based versus center-based cardiac rehabilitation. Cureus 2022 Mar;14(3):e23485. [doi: 10.7759/cureus.23485] [Medline: 35475109]
23. Antoniou V, Kapreli E, Davos CH, Batalik L, Pepera G. Safety and long-term outcomes of remote cardiac rehabilitation in coronary heart disease patients: a systematic review. Digit Health 2024;10:20552076241237661. [doi: 10.1177/20552076241237661] [Medline: 38533308]
24. Braver J, Marwick TH, Oldenburg B, Issaka A, Carrington MJ. Digital health programs to reduce readmissions in coronary artery disease: a systematic review and meta-analysis. JACC Adv 2023;2(8):100591. [doi: 10.1016/j.jacadv.2023.100591] [Medline: 38938339]
25. Thomas RJ, Beatty AL, Beckie TM, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Cardiopulm Rehabil Prev 2019 Jul;39(4):208-225. [doi: 10.1097/HCR.0000000000000447] [Medline: 31082934]
26. Nwokeji U, Spaulding EM, Shan R, et al. Health information technology use among persons with self-reported atherosclerotic cardiovascular disease: analysis of the 2011-2018 National Health Interview Survey. J Med Internet Res 2021 Aug 13;23(8):e23765. [doi: 10.2196/23765] [Medline: 34397391]
27. Mai F, Ko DG, Shan Z, Zhang D. The impact of accelerated digitization on patient portal use by underprivileged racial minority groups during COVID-19: longitudinal study. J Med Internet Res 2023 Aug 9;25:e44981. [doi: 10.2196/44981] [Medline: 37384810]
28. Wakeman M, Buckman DW, El-Toukhy S. Disparities in digital health care use in 2022. JAMA Netw Open 2025 Apr 1;8(4):e255359. [doi: 10.1001/jamanetworkopen.2025.5359] [Medline: 40244585]
29. Anawade PA, Sharma D, Gahane S. A comprehensive review on exploring the impact of telemedicine on healthcare accessibility. Cureus 2024 Mar;16(3):e55996. [doi: 10.7759/cureus.55996] [Medline: 38618307]
30. Ferrel-Yui D, Candelaria D, Pettersen TR, Gallagher R, Shi W. Uptake and implementation of cardiac telerehabilitation: a systematic review of provider and system barriers and enablers. Int J Med Inform 2024 Apr;184:105346. [doi: 10.1016/j.ijmedinf.2024.105346] [Medline: 38281451]

XSL•FO
RenderX

31. Bomtempo AP, Main E, Ghisi GL. Remote exercise engagement among individuals with cardiovascular disease: a systematic review of barriers and facilitators. J Cardiopulm Rehabil Prev 2025 Jan 1;45(1):8-19. [doi: 10.1097/HCR.0000000000000899] [Medline: 39298562]
32. Toft BS, Rodkjær L, Sørensen L, Saugbjerg MR, Bekker HL, Modrau IS. Feasibility of early digital health rehabilitation after cardiac surgery in the elderly: a qualitative study. BMC Health Serv Res 2024 Jan 22;24(1):113. [doi: 10.1186/s12913-024-10601-3] [Medline: 38254129]

Abbreviations

CABG: coronary artery bypass grafting
CR: cardiac rehabilitation
CVD: cardiovascular disease
HIT: health information technology
NSTEMI: non–ST-elevation myocardial infarction
PCCU: progressive cardiac care unit
PCI: percutaneous coronary intervention
STEMI: ST-elevation myocardial infarction
TAVR: transcatheter aortic valve replacement

Edited by J Rivers; submitted 19.Jan.2026; peer-reviewed by B Ali, SA Dabravolski; revised version received 26.Jun.2026; accepted 06.Jul.2026; published 11.Sep.2026.

Please cite as:
Patel HG, Ding J, Zhang X, Cherian A, Hyunh P, Sharma A, Avula V, Shan R, Ornstein KA, Khoury SR, Marvel FA, Mathews LM, Martin SS, Spaulding EM
Digital Health Use and Interest in Virtual Cardiac Rehabilitation Among Cardiac Inpatients: Cross-Sectional Study
JMIR Cardio 2026;10:e91767
URL: https://cardio.jmir.org/2026/1/e91767
doi:10.2196/91767

© Hamza G Patel, Jie Ding, Xinyi Zhang, Abel Cherian, Pauline Hyunh, Apurva Sharma, Vennela Avula, Rongzi Shan, Katherine A Ornstein, Shireen R Khoury, Francoise A Marvel, Lena M Mathews, Seth S Martin, Erin M Spaulding. Originally published in JMIR Cardio (https://cardio.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Postangiography Prediction of Renal Replacement Therapy in Acute Myocardial Infarction–Related Cardiogenic Shock: Least Absolute Shrinkage and Selection Operator Nomogram Development and Validation

Amin Daoulah[1], MD; Julia M Ladna[2], DO; Abdulrahman Arabi[3], MD; Ahmed Elmahrouk[1,4], MD; Shaber Seraj[2], DO; Hatem M Aloui[5], MD; Mohammed Alshehri[6], MD; Wael Almahmeed[7], MD; Prashanth Panduranga[8], MD; Nooraldaem Yousif[9], MD; Amir Lotfi[2], MD; Mohammed Qutub[10], MD; Waleed Alharbi[11], MD; Rajesh Rajan[12], MD; Mokhtar Abdirahman Kahin[13], MD; Abdullah Alenezi[14], MD; Said Al Maashani[15], MD; Taher Hassan[16], MD; Jassim Alswuaidi[3], MD; Fakhreldein Eltaieb[1], MD; Mubarak Abdulhadi Aldossari[5], MD; Mohammed Al Jarallah[12], MD; Abdelmaksoud Elganady[17,18], MD; Badr Alzahrani[19], MD; Abdulrahman Alqahtani[20], MD; Faisal Omar M Al Nasser[21], MD; Husam A Noor[9], MD; Mohamed Nabil Alama[10], PhD; Alaa Aldossari[20], MD; Sultan Al Obaikan[22], MD; Alsayed Ali Almarghany[17], MD; Omar Kanbr[23,24,25], MBBS; Youssef Elmahrouk[4], MS; Ibrahim A M Abdulhabeeb[26], MD; Mohammed Balghith[27], MD; Ahmad S Hersi[11], MD; Harvey Antony[8], MD; Adnan Fathey Hussien[13], MD; Abdulrahman Almoghairi[19], MD; Mohamed Mohammednabil A Alama[10], MD; Mohamed Ajaz Ghani[21], MD; Hazim Rahbi[5], MD; Ayman Basardah[11], MD; Bandar Alshehri[19], MD; Abdulhadi Alama[1], MD; Laura AlObaid[12,28], MD; Seraj Abualnaja[13], MD; Tarique Shahzad Chachar[9], MD; Hassan Khan[7], MD; Shahrukh Hashmani[7], MD; Ahmed A Ghonim[10], MD; Khalid Almerri[14], MD; Razan Alsofayan[20], MD; Abdelaziz M Tawengi[3], MD; Abeer M Shawky[17,18], MD; Ali Alshehri[29], MD; Amr Arafat[4,19,30], MD; Ahmed Jamjoom[1], MD, PhD

[1]King Faisal Specialist Hospital & Research Centre, Jeddah, Makkah, Saudi Arabia
[10]King Abdulaziz University, Jeddah, Saudi Arabia
[11]King Saud University, Riyadh, Saudi Arabia
[12]Amiri Hospital, Sharq, Kuwait, Kuwait
[13]International Medical Center, Jeddah, Makkah, Saudi Arabia
[14]Chest Diseases Hospital, Shuwaikh, Kuwait
[15]Sultan Qaboos Hospital, Salalah, Oman
[16]Bugshan Hospital, Jeddah, Saudi Arabia
[17]Al-Azhar University, Cairo, Egypt
[18]Dr Erfan and Bagedo General Hospital, Jeddah, Saudi Arabia
[19]Prince Sultan Cardiac Center, Riyadh, Saudi Arabia
[2]University of Massachusetts Chan Medical School, Worcester, MA, United States
[20]King Fahad Medical City, Riyadh, Saudi Arabia
[21]Department of Cardiology, Madinah Cardiac Center, Madinah, Saudi Arabia
[22]King Abdulaziz Medical City, Riyadh, Saudi Arabia
[23]Elrazi University, Khartoum, Sudan
[24]Umm al-Qura University, Makkah, Saudi Arabia
[25]Departement of Internal Medicine, Al Salama Hospital, Jeddah, Saudi Arabia
[26]Department of Cardiology, King Abdulaziz Specialist Hospital, Sakaka, Saudi Arabia
[27]King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[28]Royal College of Surgeons in Ireland, Dublin, Ireland
[29]King Khalid University, Abha, Saudi Arabia
[3]Hamad Medical Corporation, Doha, Qatar
[30]Research and Innovation Institute, Ministry of Defense Health Services, Riyadh, Saudi Arabia
[4]Tanta University, Tanta, Gharbiya, Egypt
[5]King Saud Medical City, Riyadh, Saudi Arabia
[6]Armed Forces Hospital Southern Region, Khamis Mushait, Saudi Arabia
[7]Heart & Vascular Institute, Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates
[8]Department of Cardiology, National Heart Center, Royal Hospital, Muscat, Oman
[9]Department of Cardiology, Mohammed Bin Khalifa Specialist Cardiac Center, Awali, Bahrain

XSL•FO
RenderX

Corresponding Author:
Ahmed Elmahrouk, MD
King Faisal Specialist Hospital & Research Centre, Jeddah, Makkah, Saudi Arabia

Abstract

Background: Acute kidney injury critically impacts outcomes in cardiogenic shock secondary to acute myocardial infarction (CS-AMI). Acute kidney injury is one of the strongest independent predictors of in-hospital mortality in CS-AMI. Despite evidence that early renal replacement therapy (RRT) initiation improves survival, comprehensive prediction models for RRT in this population remain lacking.

Objective: This study aimed to develop and internally validate a Least Absolute Shrinkage and Selection Operator (LASSO) regression–based prediction model and clinical nomogram for in-hospital RRT in patients with CS-AMI.

Methods: This multicenter retrospective cohort study included 1431 patients with CS-AMI from the Gulf Cardiogenic Shock (Gulf-CS) registry across 13 centers in 6 Gulf countries (2020 - 2022). LASSO logistic regression was applied to a training set (1071/1431, 80%) to select baseline predictors of RRT; performance was evaluated on a held-out testing set (268/1431, 20%). Internal validation included 10-fold cross-validation and bootstrapping (1000 iterations). Cluster-robust SEs accounted for center effects. The model was compared to a parsimonious model (age+creatinine clearance), and a clinical nomogram was developed.

Results: Of 1431 patients, 190 (13.3%) required RRT. Patients requiring RRT were significantly older (mean 64.17, SD 12.14 y vs mean 59.75, SD 11.77 y; *P*<.001), with higher prevalences of diabetes mellitus (72.1% vs 61.9%; *P*=.008), peripheral arterial disease (11.6% vs 3.7%; *P*<.001), and prior cerebrovascular accident (11.1% vs 5.7%; *P*=.005). The RRT group had lower creatinine clearance (46 vs 72 mL/min; *P*<.001), higher baseline lactate (2.7 vs 2.1 mmol/L; *P*<.001), and more advanced Society for Cardiovascular Angiography and Interventions (SCAI) shock stages (stages D and E: 90.5% vs 64.9%; *P*<.001). LASSO selected 15 baseline predictors. The model achieved an area under the receiver operating characteristic curve (AUC) of 0.714 on the testing set, significantly outperforming the parsimonious model (AUC: 0.624; *P*<.001). Bootstrap-corrected AUC was 0.745 (95% CI 0.730 - 0.756). In-hospital mortality was markedly higher in the RRT group (75.8% vs 38.8%; *P*<.001), with longer hospital stay (10 vs 6 d; *P*<.001), more major bleeding (16.8% vs 7.3%; *P*<.001), and cerebrovascular accidents (11.1% vs 4.9%; *P*=.001).

Conclusions: We have developed and internally validated a robust 15-variable nomogram (Gulf-CS-Nomogram) that accurately predicts the need for RRT in patients with CS-AMI using baseline data intended for use after coronary angiography. This tool may facilitate early nephrology consultation and timely RRT initiation to improve outcomes.

(JMIR Cardio 2026;10:e79678) doi:10.2196/79678

KEYWORDS

myocardial infarction; acute myocardial infarction; cardiogenic shock; renal replacement therapy; prediction model; Least Absolute Shrinkage and Selection Operator; LASSO; machine learning

Introduction

Cardiogenic shock (CS) is an end-organ damage and tissue hypoxia caused by cardiac dysfunction, with acute myocardial infarction (AMI) being the most common underlying cause. AMI leads to myocardial necrosis, loss of ventricular function, and reduced cardiac output, making CS a leading cause of in-hospital mortality in patients with AMI [1]. Within the United States, the incidence of CS in patients with AMI ranges from 5% to 10%, and despite advances in treatment, the incidence has been increasing since 2000 [2,3]. Mortality remains high, with 30-day mortality rates of approximately 40% and 1-year mortality rates approaching 50%. Survivors of CS often experience significant morbidity, including high rates of readmission and the need for ongoing supportive care [1].

Acute kidney injury (AKI) is a common and severe complication in patients with cardiogenic shock secondary to acute myocardial infarction (CS-AMI). In this context, the primary driver of AKI is reduced cardiac output, leading to decreased renal perfusion and increased renal venous pressure [4,5]. Additional contributors include microvascular injury and hemodynamic instability [6]. Studies report AKI incidence rates ranging from 35% to 55% in patients with CS-AMI, with approximately 13% requiring renal replacement therapy (RRT) [6,7]. AKI is associated with worse outcomes, including increased in-hospital mortality, longer lengths of stay, and higher health care costs. Notably, AKI is the strongest independent predictor of in-hospital mortality in patients with CS-AMI [8,9]. Lauridsen et al [7] reported that AKI requiring RRT in patients with CS-AMI was associated with a 62% in-hospital mortality rate, and survivors faced a significantly greater risk of chronic dialysis than patients with non-AKI (11% vs 1%).

Several studies have highlighted the potential benefits of early initiation of RRT in critically ill patients. For example, the Early versus Late Initiation of RRT in Critically Ill Patients with Acute Kidney Injury (ELAIN) trial demonstrated that early RRT initiation significantly reduced 90-day mortality compared with

XSL•FO
RenderX

delayed RRT initiation (39.3% vs 54.7%) and improved renal recovery [10]. Similarly, a meta-analysis by Zou et al [11] revealed that early RRT reduced 30-day mortality and shortened the length of stay in postcardiac surgery patients with AKI.

These findings highlight the critical impact of AKI on prognosis and early intervention on mortality in patients with CS-AMI. Few studies have identified risk factors or predictors for RRT in this population. Some known predictors include a reduced glomerular filtration rate at admission, elevated central venous pressure, decreased mixed venous oxygen saturation, increased age, increased comorbidity burden, and preexisting chronic kidney disease [9,12]. Comprehensive models predicting the need for RRT in patients with CS-AMI are still needed. In this study, we aimed to predict RRT in this population via Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression. Predicting the need for RRT could help reduce mortality and improve outcomes.

Methods

Study Population

The Gulf Cardiogenic Shock (Gulf-CS) registry is a multicenter retrospective registry of CS secondary to angiographically confirmed myocardial infarction [13]. The registry included 1513 patients from 13 tertiary referral centers in 6 Gulf countries (Saudi Arabia, Qatar, Oman, the United Arab Emirates, Kuwait, and Bahrain) between January 2020 and December 2022. Data on baseline demographics, comorbidities, number of pressors or inotropes, angiographic findings, hemodynamic measurements, and mechanical circulatory devices were collected.

Patients with baseline creatinine clearance <15 mL/minute and those with chronic kidney disease requiring dialysis at presentation were excluded. For this analysis, we divided the patients into 2 groups based on their need for RRT: those who did not require RRT (n=1241) and those who required RRT (n=190). We compared the two groups' baseline variables, angiographic findings, treatment, and in-hospital outcomes. RRT used in this study was hemodialysis (n=54), peritoneal dialysis (n=1), and continuous venovenous hemodialysis (n=135).

Study Outcomes

Hospital outcomes were compared between patients with and without RRT. The primary outcome was in-hospital all-cause mortality. The secondary outcomes included revascularization, cerebrovascular accident (CVA), the occurrence of major or minor bleeding, and the duration of hospital stay.

Study Definitions

CS was defined according to the Standardized Definitions for Cardiogenic Shock Research and Mechanical Circulatory Support Devices as follows: (1) systolic blood pressure <90 mm Hg for ≥30 minutes or the need for vasopressors, inotropes, or mechanical circulatory support (MCS) to maintain a systolic blood pressure ≥90 mm Hg and (2) evidence of tissue hypoperfusion [14]. The time of shock onset is the time when CS is first diagnosed. CVAs include stroke (ischemic or hemorrhagic) and transient ischemic attack. Bleeding events were defined according to the Bleeding Academic Research Consortium (BARC) classification [15]. Creatinine clearance was calculated using the Cockcroft-Gault equation: CrCl=[(140 – age)×weight (kg) × (0.85 if female)]/[72×serum creatinine (mg/dL)]. We chose this method over estimated glomerular filtration rate equations because of its established use in the CS literature, its incorporation of body weight (relevant to acute volume shifts), and its direct applicability to medication dosing decisions. Right ventricular (RV) dysfunction was defined by echocardiographic criteria: tricuspid annular plane systolic excursion <17 mm, S' wave <12 cm/s, and tricuspid annular plane systolic excursion to pulmonary artery systolic pressure ratio <0.34 [16]. Baseline data, including laboratory and echocardiographic data, were measured at the time of hospital presentation. The number of diseased vessels was defined as the total count of coronary arterial branches with angiographically significant stenosis (≥50% luminal obstruction), including major epicardial vessels and their principal branches (left anterior descending artery, diagonal branches, obtuse marginal branches, ramus intermedius, right coronary artery, posterior descending artery, and posterolateral branches), yielding a possible range of 1 to 6 in this cohort. Details of the Gulf-CS registry and variable definitions have already been published [13].

Ethical Considerations

The Institutional Review Board (IRB) of King Faisal Specialist Hospital and Research Center in Jeddah approved the study on January 16, 2023 (IRB 2022 - 90: Gulf-CS Registry). Approval was obtained from each center's ethics committee for all study sites. The IRB waived the need for patient consent due to the study's observational nature. Patients' privacy was maintained throughout the study, and deidentified data were shared. No compensation was provided to participants.

Statistical Analysis

Statistical analysis was conducted with Stata 18 (StataCorp) and Python 3.11 (Python Software Foundation). Continuous variables were summarized using means (SDs) or medians (Q1-Q3) according to the normal distribution. Categorical data are reported as numbers (%). Comparisons between the two groups were made using *t* tests, Mann-Whitney tests, chi-square tests, or Fisher exact tests, as appropriate. Missing data were handled using complete case analysis.

Model Development and Variable Selection

The dataset was randomly partitioned into a training set (80%) and a testing set (20%), stratified by the primary outcome of in-hospital RRT. We used the LASSO logistic regression on the training set to identify the most parsimonious set of baseline predictors. The LASSO method applies a penalty that shrinks the coefficients of less important variables toward zero, effectively performing automated variable selection. The tuning parameter (lambda) was selected via 10-fold cross-validation to maximize the area under the receiver operating characteristic curve (AUC).

XSL•FO
RenderX

Model Performance and Validation

The model's discrimination was assessed on the held-out test set using the AUC. We performed two forms of internal validation:

1. 10-fold cross-validation on the training set.
2. Bootstrapping with 1000 resamples on the training set to generate a bias-corrected 95% CI for the AUC.

Model calibration was assessed visually by plotting the observed versus predicted probabilities on the testing set.

Model Calibration

Model calibration was assessed visually by plotting the observed versus predicted probabilities on the testing set. To improve calibration, we applied Platt scaling (also known as sigmoid calibration) to the LASSO model predictions. Platt scaling fits a logistic regression model to the predicted probabilities from the LASSO model, effectively recalibrating the predictions to better match the observed event rates. This post hoc calibration was performed using 5-fold cross-validation on the training set, and the calibrated model was then applied to the testing set. Calibration performance was evaluated both before and after Platt scaling using calibration plots.

Nomogram Development

A clinical nomogram was constructed to provide individualized, point-based estimates of the probability of requiring RRT. The nomogram was derived from the LASSO logistic regression model described above, using the penalized coefficients of all 15 candidate variables retained at the optimal regularization parameter. To facilitate clinical application, the standardized LASSO coefficients were backtransformed to their original measurement scales by dividing each coefficient by the corresponding variable's SD, with continuous variables expressed in their natural clinical units. For each predictor, a point score proportional to its backtransformed coefficient and clinical range was assigned relative to the variable with the largest absolute contribution; the sum of individual point scores maps to a predicted probability of RRT via the logistic function. The nomogram is intended as a bedside decision-support tool to stratify patients with CS-AMI by their risk of acute kidney injury requiring RRT using variables available at or shortly after admission.

Addressing Center Effect and Model Comparison

To account for potential clustering of outcomes by institution, we fitted a final logistic regression model using the LASSO-selected variables with cluster-robust SEs for each center. Continuous variables were entered into the final model as unstandardized values to yield clinically interpretable odds ratios (ORs). The Society for Cardiovascular Angiography and Interventions (SCAI) shock stage was treated as a categorical variable with stage B as the reference category, as this approach more accurately captures the nonlinear risk transitions between stages. To demonstrate the clinical utility of our model, we compared its performance on the test set against a parsimonious model that included only age and baseline creatinine clearance.

Multicollinearity

We assessed multicollinearity among the predictors in the final model using the variance inflation factor (VIF). A VIF >5 was considered indicative of significant multicollinearity. All included features had VIF <1.5.

Results

Baseline Characteristics

The baseline characteristics of the patients in our study revealed several demographic and clinical differences between those who required RRT and those who did not. The cohort's mean age was 60.34 (SD 12.14) years, with patients requiring RRT significantly older (mean 64.17, SD 12.14 y) than those who did not (mean 59.75, SD 11.77 y; P<.001). Males comprised most of the cohort (1162/1431, 81.20%), with no significant difference between the 2 groups (P=.46). Patients who required RRT had a greater prevalence of comorbid conditions, including diabetes mellitus (72.11% vs 61.89%; P=.008), peripheral arterial disease (PAD; 11.58% vs 3.71%; P<.001), a history of CVA (11.05% vs 5.72%; P=.005), and congestive heart failure (15.79% vs 10.64%; P=.04). Other comorbidities, such as dyslipidemia, hypertension, previous myocardial infarction, and history of coronary interventions (percutaneous coronary intervention [PCI] and coronary artery bypass grafting [CABG]), were similar between the groups. Patients requiring RRT presented with more severe disease, as evidenced by higher rates of anterior ST-elevation myocardial infarction (STEMI) and cardiac arrest at presentation. Additionally, RV dysfunction was more prevalent among patients requiring renal support (23.28% vs 15.76%; P=.01), further highlighting the increased severity of their clinical presentation.

Laboratory investigations revealed lower hemoglobin, creatinine clearance, sodium bicarbonate levels, and higher potassium levels in patients requiring RRT. Patients who required RRT had a higher SYNTAX score, a greater number of vessels with significant coronary disease, and an advanced SCAI stage than patients who did not require RRT (Table 1).

XSL•FO
RenderX

Table . Comparison of the baseline data between patients who did and did not need renal replacement therapy.

Variables	Total (N=1431)	No renal replacement therapy (n=1241)	Renal replacement therapy (n=190)	*P* value
Age (years), mean (SD)	60.34 (12.14)	59.75 (11.77)	64.17 (12.14)	<.001
Males, n (%)	1162 (81.20)	1004 (80.90)	158 (83.16)	.46
BMI (kg/m^2), median (IQR)	27 (24 - 30)	27 (24 - 30)	27 (25 - 29)	.58
Diabetes mellitus, n (%)	905 (63.24)	768 (61.89)	137 (72.11)	.008
Dyslipidemia, n (%)	811 (56.67)	701 (56.49)	110 (57.89)	.72
Hypertension, n (%)	891 (62.26)	761 (61.32)	130 (68.42)	.60
Previous myocardial infarction, n (%)	320 (22.36)	275 (22.16)	45 (23.68)	.64
History of PCI[a], n (%)	283 (18.78)	245 (19.74)	38 (20)	.93
History of CABG[b], n (%)	47 (3.28)	39 (3.14)	8 (4.21)	.44
Peripheral arterial disease, n (%)	68 (4.75)	46 (3.71)	22 (11.58)	<.001
History of CVA[c], n (%)	92 (6.43)	71 (5.72)	21 (11.05)	.005
History of CHF[d], n (%)	162 (11.32)	132 (10.64)	30 (15.79)	.04
COPD[e], n (%)	47 (3.28)	43 (3.46)	4 (2.11)	.33
Presentation, n (%)				
STEMI[f]	1061 (74.14)	948 (76.39)	113 (59.47)	<.001
NSTEMI[g]	370 (25.86)	293 (23.61)	77 (40.53)	<.001
Cardiac arrest	789 (55.14)	657 (52.94)	132 (69.47)	<.001
Tachyarrhythmia	450 (31.49)	381 (30.75)	69 (36.32)	.12
Bradyarrhythmia	202 (14.12)	169 (13.62)	33 (17.37)	.16
ECG[h], n (%)				
Location of STEMI				<.001
NSTEMI	370 (25.86)	293 (23.61)	77 (40.53)	
Anterior	689 (48.15)	611 (49.23)	78 (41.05)	
Lateral	321 (22.43)	290 (23.37)	31 (16.32)	
Inferior	51 (3.56)	47 (3.79)	4 (2.11)	
Echocardiography				
Ejection fraction, median (IQR)	32 (23 - 39)	32 (23 - 39)	33 (24 - 38)	.13
Right ventricular dysfunction, n (%)	237 (16.76)	193 (15.76)	44 (23.28)	.01
PASP[i] (mm Hg), median (IQR)	35 (30 - 40)	35 (30 - 40)	35 (31 - 40)	.16
Laboratory data, median (IQR)				
Hemoglobin (g/dL)	13.3 (11.7 - 14.9)	13.5 (12-15)	12.3 (10.9 - 14)	<.001
WBCs[j] (10^9L)	12.4 (9.8 - 15.8)	12.4 (9.8 - 15.3)	13 (10 - 16.4)	.20
Potassium (mmol/L)	4.2 (3.8 - 4.6)	4.2 (3.8 - 4.6)	4.4 (3.98 - 4.8)	.001
Creatinine clearance (mL/min)	68 (43 - 95)	72 (47 - 97)	46 (33 - 74)	<.001
Baseline lactate (mmol/L)	2.2 (1.3 - 3.8)	2.1 (1.2 - 3.5)	2.7 (1.9 - 4.6)	<.001

XSL•FO
RenderX

Variables	Total (N=1431)	No renal replacement therapy (n=1241)	Renal replacement therapy (n=190)	*P* value
HCO_3 (mEq/L)	20 (17 - 23)	20 (17 - 23)	18 (17 - 22)	<.001
Angiography, median (IQR)				
SYNTAX score	25 (19 - 32)	25 (19 - 32)	28 (22 - 33)	<.001
Vessels with significant disease	3 (2-3)	1 (2-3)	3 (2-4)	<.001
SCAI[k] stage, n (%)				<.001
B	91 (6.36)	90 (7.25)	1 (0.53)	
C	363 (25.37)	346 (27.88)	17 (8.95)	
D	704 (49.20)	596 (48.03)	108 (56.84)	
E	273 (19.08)	209 (16.84)	64 (33.68)	

[a]PCI: percutaneous coronary intervention.
[b]CABG: coronary artery bypass grafting.
[c]CVA: cerebrovascular accident.
[d]CHF: congestive heart failure.
[e]COPD: chronic obstructive pulmonary disease.
[f]STEMI: ST-elevation myocardial infarction.
[g]NSTEMI: non–ST-elevation myocardial infarction.
[h]ECG: electrocardiogram.
[i]PASP: pulmonary artery systolic pressure.
[j]WBC: white blood cell.
[k]SCAI: Society for Cardiovascular Angiography and Interventions.

CS Management

Inotropic support, MCS, and mechanical ventilation were used more frequently in patients who needed renal support. CABG was performed more frequently in patients requiring RRT, and PCI was more common in patients without RRT (Table 2).

XSL•FO
RenderX

Table . Comparison of the management data between patients who did and did not need renal replacement therapy.

Variables	Total (N=1431)	No renal replacement therapy (n=1241)	Renal replacement therapy (n=190)	*P* value
Pressor or inotropic support, n (%)	1294 (90.43)	1110 (89.44)	184 (96.84)	.001
Mechanical circulatory support, n (%)	660 (46.12)	529 (42.63)	131 (68.95)	<.001
Intraaortic balloon	503 (33.25)	399 (30.95)	104 (46.43)	
ECMO[a]	27 (1.78)	23 (1.78)	4 (1.79)	
Impella	31 (2.05)	26 (2.02)	5 (2.23)	
IABP[b]+ ECMO	95 (6.28)	71 (5.51)	24 (10.71)	
IABP+ Impella	12 (0.79)	8 (0.62)	4 (1.79)	
ECMO+ Impella	10 (0.66)	6 (0.47)	4 (1.79)	
IABP+ ECMO+Impella	15 (0.99)	11 (0.85)	4 (1.79)	
Mechanical ventilation, n (%)	878 (61.36)	702 (56.57)	176 (92.63)	<.001
Treatment, n (%)				.007
Medical	137 (9.57)	118 (9.51)	19 (10)	
PCI[c]	1174 (82.04)	1030 (83)	144 (75.79)	
CABG[d]	120 (8.39)	93 (7.49)	27 (14.21)	

[a]ECMO: extracorporeal membrane oxygenation.

[b]IABP: intraaortic balloon pump.

[c]PCI: percutaneous coronary intervention.

[d]CABG: coronary artery bypass grafting.

Hospital Outcomes

Hospital mortality (76% vs 39%; *P*<.001) and CVA (11% vs 5%, *P*=.001) were significantly greater in patients with RRT. Hospitalization duration was prolonged in patients with RRT (10 vs 6 d; *P*<.001). No differences were found in revascularizations or minor bleeding between the groups (Table 3).

Table . Comparison of hospital outcomes between patients who did and did not need renal replacement therapy.

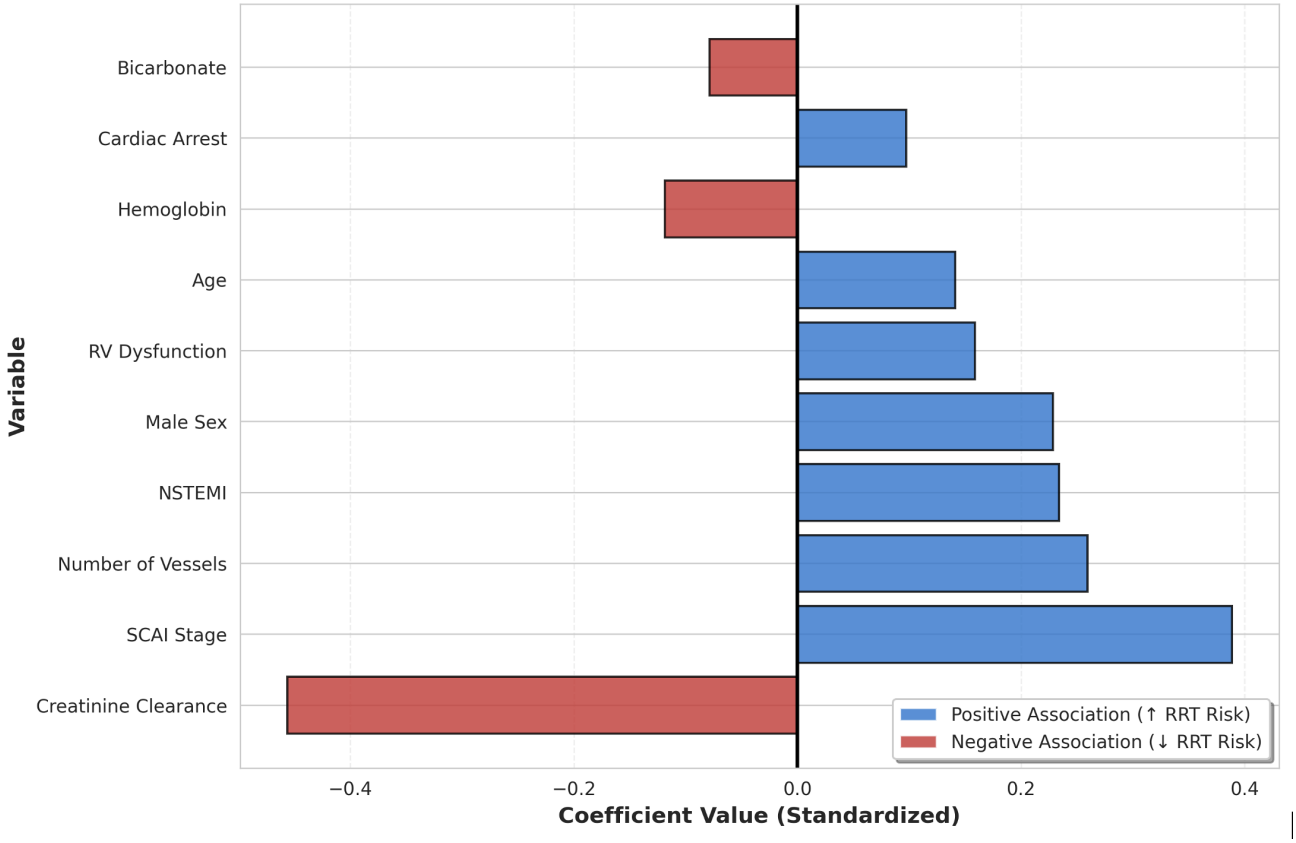
Hospital outcomes	Total (N=1431)	No renal replacement therapy (n=1241)	Renal replacement therapy (n=190)	*P* value
Hospital mortality, n (%)	625 (43.68)	481 (38.76)	144 (75.79)	<.001
Target lesion revascularization, n (%)	57 (3.98)	49 (3.95)	8 (4.21)	.86
Target vessel revascularization, n (%)	43 (3)	37 (2.98)	6 (3.16)	.89
Cerebrovascular accident, n (%)	82 (5.73)	61 (4.92)	21 (11.05)	.001
Major bleeding, n (%)	123 (8.60)	91 (7.33)	32 (16.84)	<.001
Minor bleeding, n (%)	27 (1.89)	23 (1.85)	4 (2.11)	.77
Duration of hospital stay (d), median (IQR)	6 (3-13)	6 (2-12)	10 (4 - 22)	<.001

LASSO Model Development

The LASSO model included 1072 in the training dataset and 268 in the test dataset (a total of 1340), while 91 patients were dropped due to missing baseline data. From the baseline variables, the LASSO regression model selected a final set of 15 predictors for RRT. The selected variables and their corresponding coefficients are shown in Figure 1. The 3 strongest predictors were lower creatinine clearance, higher SCAI shock stage, and a greater number of diseased coronary vessels (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Least Absolute Shrinkage and Selection Operator–selected variable coefficients for renal replacement therapy prediction. Top 10 baseline predictors ranked by absolute coefficient value. Blue bars: positive association with renal replacement therapy risk; red bars: negative association with renal replacement therapy risk. All coefficients are standardized. NSTEMI: non–ST-elevation myocardial infarction; RRT: renal replacement therapy; RV: right ventricular; SCAI: Society for Cardiovascular Angiography and Interventions.

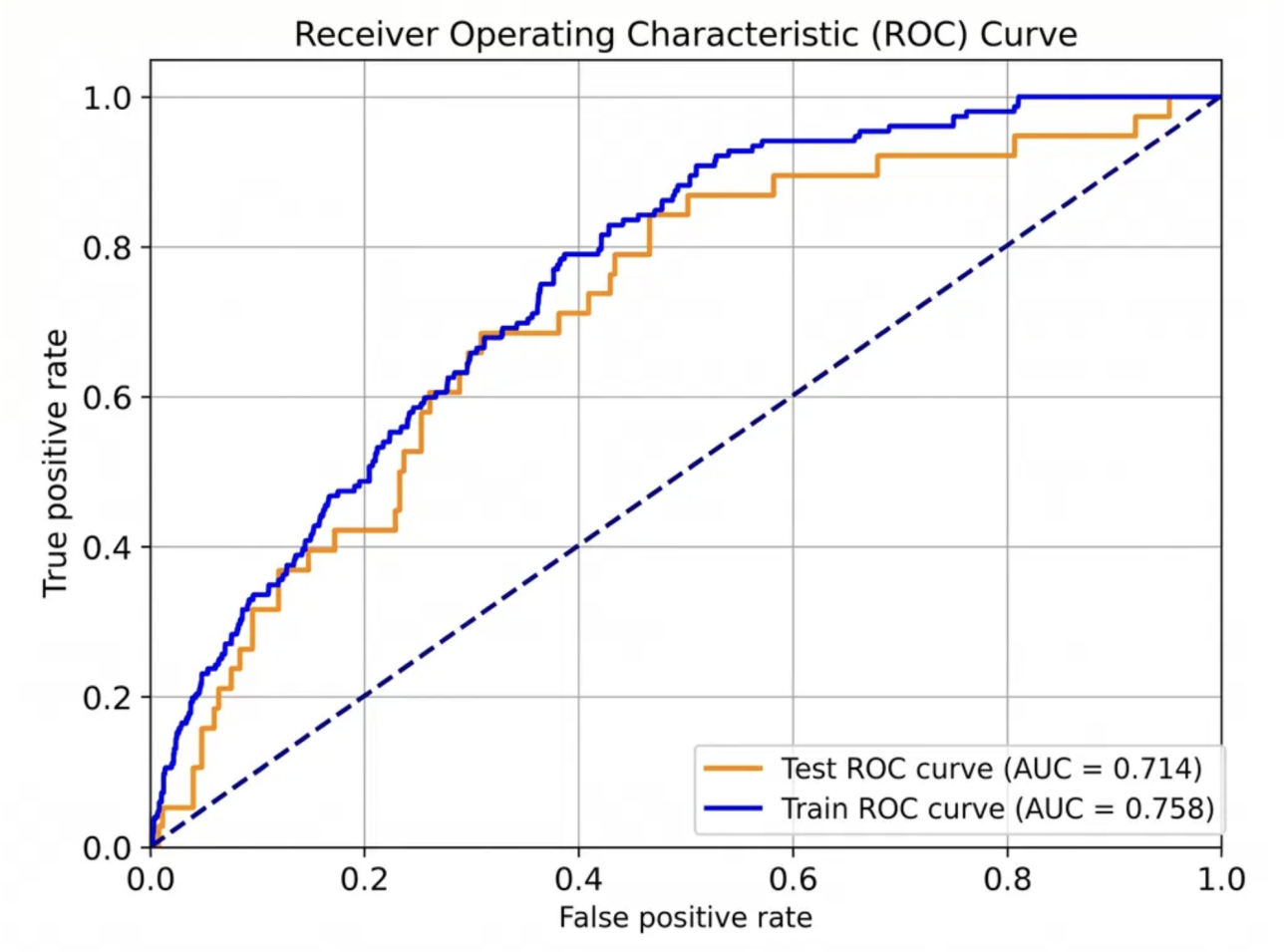


Model Performance

The final model demonstrated good discrimination on the held-out testing set, with an AUC of 0.714 (Figure 2). This was significantly better than the parsimonious model containing only age and creatinine clearance, which had an AUC of 0.624 (P<.001 for the difference; Figure 3). The model was well-calibrated, as shown in the calibration plot (Figure S1 in Multimedia Appendix 2).

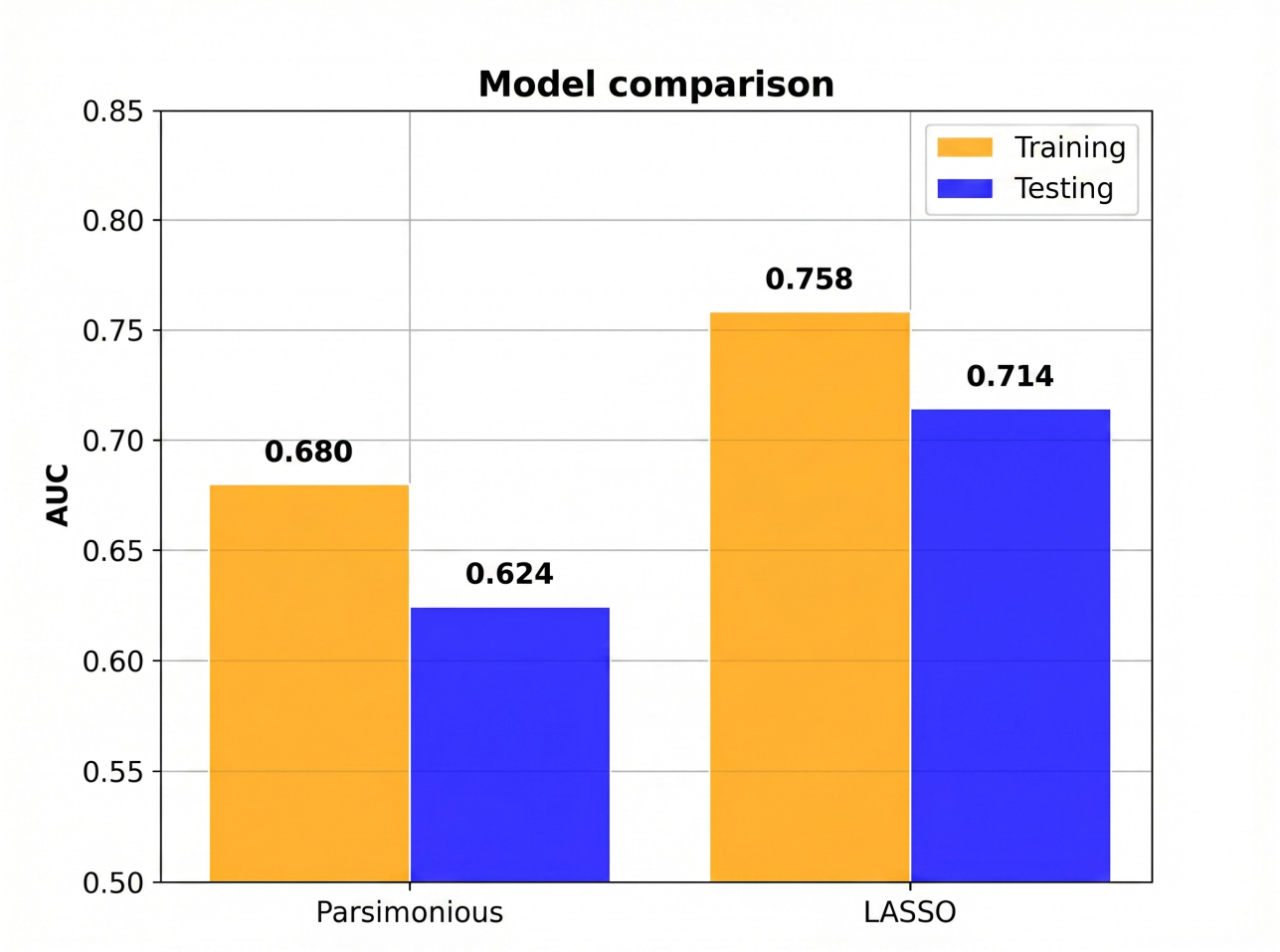
XSL•FO
RenderX

Figure 2. Receiver operating characteristic curves for the Least Absolute Shrinkage and Selection Operator prediction model. Blue curve: training set (AUC=0.758, n=1072); orange curve: testing set (AUC=0.714, n=268). Dashed line: reference (AUC=0.50). The model demonstrated good discrimination with minimal overfitting. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic.



XSL•FO
RenderX

Figure 3. Model performance comparison. Orange bars: training set (n=1072); blue bars: testing set (n=268). The 15-variable Least Absolute Shrinkage and Selection Operator model (AUC=0.714) significantly outperformed the 2-variable parsimonious model containing only age and creatinine clearance (AUC=0.624), with an improvement of +0.090 AUC points (*P*<.001). AUC: area under the receiver operating characteristic curve; LASSO: Least Absolute Shrinkage and Selection Operator.



Internal Validation

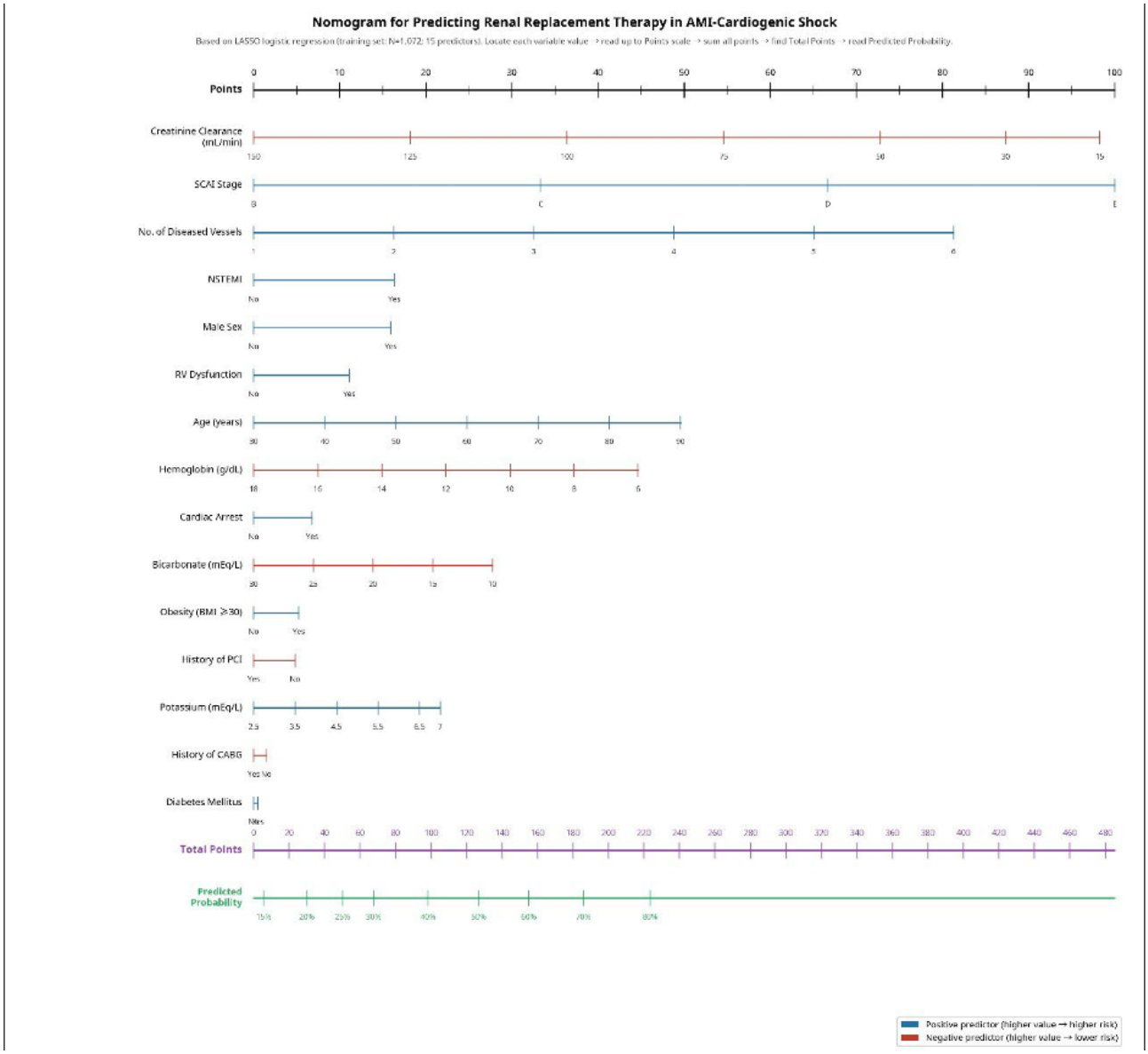
The model's performance was confirmed through 2 internal validation methods. A 10-fold cross-validation on the training set yielded a mean AUC of 0.723 (SD 0.050). Furthermore, bootstrapping with 1000 iterations produced a bias-corrected AUC of 0.745 (95% CI 0.730 - 0.756), indicating robust and stable performance (Figure S2 in Multimedia Appendix 3).

Nomogram

A nomogram was developed based on the final LASSO model (Figure 4). A nomogram can be used post angiography immediately to estimate an individual patient's probability of requiring in-hospital RRT by summing the points assigned to each predictor variable and reading the corresponding probability from the total points scale.

XSL•FO
RenderX

Figure 4. Gulf Cardiogenic Shock Nomogram for predicting in-hospital renal replacement therapy in patients with acute myocardial infarction complicated by cardiogenic shock. The nomogram was derived from the Least Absolute Shrinkage and Selection Operator logistic regression model. For each variable, the patient's value is located on the corresponding row and projected vertically to the points scale (0 - 100). Individual points are summed to obtain total points, which are then projected to the predicted probability of renal replacement therapy scale. Blue axes indicate positive associations with renal replacement therapy risk; red axes (creatinine clearance, hemoglobin, bicarbonate, history of percutaneous coronary intervention, and history of coronary artery bypass grafting) indicate negative associations, with axes running right to left such that lower values (higher risk) correspond to more points. CABG: coronary artery bypass grafting; NSTEMI: non–ST-segment elevation myocardial infarction; PCI: percutaneous coronary intervention; RV: right ventricular; SCAI: Society for Cardiovascular Angiography and Interventions.



Multivariable Analysis With Cluster-Robust SEs

In the final logistic regression model accounting for center-level clustering, several predictors remained significantly associated with the need for RRT. These included history of PCI (OR 0.649; *P*=.007), non–ST-elevation myocardial infarction (NSTEMI) presentation (OR 1.756; *P*=.02), SCAI stage D (OR 10.859; *P*=.02) and E (OR 14.613; *P*=.005), lower hemoglobin (OR 0.892; *P*=.03), and increased number of diseased vessels (OR 1.351; *P*=.001; Table 4). No significant multicollinearity was detected among the predictors (all VIF <1.5).

XSL•FO
RenderX

Table . Multivariable logistic regression with cluster-robust SEs for prediction of in-hospital renal replacement therapy. Logistic regression model fitted across multiple centers (n=1340). Standard errors are adjusted for clustering by center. Society for Cardiovascular Angiography and Interventions stage B is the reference category.

Variable	Odds ratio (95% CI)	*P* value
Age (per 10-y increase)	1.178 (0.996 - 1.393)	.06
Sex (male)	1.925 (0.923 - 4.012)	.08
Obesity (BMI ≥30)	1.209 (0.922 - 1.586)	.17
Diabetes mellitus	1.031 (0.773 - 1.375)	.84
History of PCI[a]	0.649 (0.473 - 0.891)	.007
History of CABG[b]	0.838 (0.343 - 2.045)	.70
NSTEMI[c] (vs STEMI[d])	1.756 (1.081 - 2.851)	.02
Cardiac arrest	1.613 (0.938 - 2.772)	.08
RV[e] dysfunction	1.506 (0.852 - 2.665)	.16
SCAI[f] stage C (vs B)	4.740 (0.596 - 37.675)	.14
SCAI stage D (vs B)	10.859 (1.439 - 81.924)	.02
SCAI stage E (vs B)	14.613 (2.232 - 95.676)	.005
Hemoglobin (per 1 g/dL)	0.892 (0.804 - 0.989)	.03
Potassium (per 1 mEq/L)	1.186 (0.934 - 1.506)	.16
Bicarbonate (per 1 mEq/L)	0.975 (0.927 - 1.025)	.32
Creatinine clearance (per 1 mL/min)	0.996 (0.986 - 1.006)	.39
Number of diseased vessels	1.351 (1.181 - 1.545)	.001

[a]PCI: percutaneous coronary intervention.

[b]CABG: coronary artery bypass grafting.

[c]NSTEMI: non–ST-segment elevation myocardial infarction.

[d]STEMI: ST-segment elevation myocardial infarction.

[e]RV: right ventricular.

[f]SCAI: Society for Cardiovascular Angiography and Interventions.

Discussion

Principal Findings

This study used data from the Gulf-CS registry, the first large multinational study focused on patients with CS-AMI from the Gulf region. This study explored hospital and short-term outcomes of patients with CS-AMI across 6 Gulf countries, using the novel CS staging system developed by the SCAI Cardiogenic Shock Working Group (SCAI-CSWG). We used LASSO regression to predict the need for RRT in this high-risk population. Our analysis identified several predictors of RRT, including reduced creatinine clearance, advanced SCAI stage, and significant coronary artery disease affecting multiple vessels.

Renal support was required in 13.28% of patients (190/1431), and those requiring RRT had a significantly worse prognosis. This finding is consistent with previous studies showing that the RRT incidence in patients with CS-AMI is approximately 13% [7,17]. Patients who required renal support were notably older, with a mean age of 64.17 (SD 12.14) years, whereas those who did not require RRT were 59.75 (SD 11.77) years (*P*<.001), which is also consistent with the literature [9,18]. Comorbidities were also more prevalent in the renal support group, with higher rates of diabetes mellitus (72.11% vs 61.89%; *P*=.008), PAD (11.58% vs 3.71%; *P*<.001), previous CVA (11.05% vs 5.72%; *P*=.005), and congestive heart failure (15.79% vs 10.64%; *P*=.037). The literature also highlights a high comorbidity burden associated with increased renal support needs [19]. Patients presenting with NSTEMI and those experiencing cardiac arrest were also more likely to require RRT. Specifically, 40.53% (77/190) of patients requiring RRT had NSTEMI, whereas 23.61% (293/1241) of patients in the no-renal-support group did (*P*<.001). Cardiac arrest was present in 69.47% (132/190) of patients in the RRT group, which was significantly greater than the 52.94% reported in patients who did not require RRT (*P*<.001). These findings highlight the association between more severe initial presentations and the subsequent need for RRT. In a study by Jeppesen et al [18], 33.5% of patients who experienced out-of-hospital cardiac arrest developed AKI, and 10.1% required RRT.

Regarding echocardiographic findings, RV dysfunction was significantly more prevalent in the renal support group, suggesting that RV failure and venous congestion contribute to worsening renal function. This finding is corroborated by Jain and associates, who demonstrated that RV dysfunction is

XSL•FO
RenderX

common in patients with CS and is associated with increased in-hospital mortality [20]. RV failure is associated with poor outcomes and often requires MCS [21]. Furthermore, baseline laboratory marker levels, such as creatinine clearance and hemoglobin, potassium, and lactate levels, were significantly lower in patients who required RRT. The median creatinine clearance was markedly lower in the RRT group (46 mL/min vs 72 mL/min; *P*<.001), and the baseline lactate level was greater (2.7 mmol/L vs 2.1 mmol/L; *P*<.001), indicating more pronounced tissue hypoperfusion. This finding is consistent with other studies that associated worse laboratory abnormalities with renal support [22].

Finally, angiographic findings such as the SYNTAX score, number of major vessels involved, and SCAI stage were also obtained. The data suggested that the SYNTAX score was greater in the renal support group, with more significant coronary disease in the RRT group. This finding is consistent with prior findings that lesion complexity increases with declining renal function [23]. Significant differences in SCAI stages, with more severe stages (D and E), were also observed in the RRT group (*P*<.001).

Mortality and Outcomes

The need for RRT was associated with a significantly higher in-hospital mortality rate. Patients requiring RRT had a mortality rate of 75.79%, whereas 38.76% (144/190 and 481/1241, respectively) of those who did not require RRT (*P*<.001). This finding is similar to that of a study performed by Lauridsen et al [7], which revealed that RRT in patients with CS-AMI was associated with worse mortality (62% for patients with AKI-RRT vs 36% for patients with non-AKI-RRT), along with increased long-term mortality at 5 years. This significant difference highlights the role of renal failure as a major determinant of poor outcomes in patients with CS-AMI. In addition to higher mortality, patients requiring renal support had significantly longer hospital stays (median of 10 d, IQR 4 - 22 vs 6 d; *P*<.001) and higher rates of major bleeding complications (16.84% vs 7.33%; *P*<.001). Similarly, external studies have shown a longer length of stay; for example, Vallabhajosula et al [9] reported that patients with CS-AMI requiring hemodialysis had a significantly longer length of stay (mean 18, SD 19 d vs mean 9, SD 10 d). These adverse outcomes likely reflect the increased severity of illness in patients requiring RRT, as well as the risks associated with invasive procedures and anticoagulation in this population.

Other clinical outcomes, such as stroke and bleeding complications, were also more frequent in patients who required RRT. The incidence of cerebrovascular accidents was 11.05% in the renal support group, whereas it was 4.92% in those not requiring RRT (*P*=.001). Major bleeding occurred in 16.84% (32/190) of patients receiving renal support, which was more than double the rate in those who did not receive RRT (91/1241, 7.33%; *P*<.001). Freund and colleagues reported from a subanalysis of the CULPRIT-SHOCK trial that bleeding events were common in patients with infarct-related CS (21.5%), particularly in those with MCS, which is often used with renal support [24]. These complications likely contribute to the higher mortality and morbidity in this group, highlighting the complexity of managing these critically ill patients.

LASSO Regression

LASSO regression is used in this study to handle the large number of potential predictors and avoid overfitting. By penalizing the absolute size of the regression coefficients, LASSO effectively selects the most relevant variables, making it an ideal tool for this analysis. LASSO identified several predictors of renal support in patients with CS-AMI. Key predictors included lower creatinine clearance, advanced SCAI stage, and multivessel coronary artery disease. Previous studies, such as that by Marenzi et al [8], have established that mechanical ventilation (OR 22.6, 95% CI 14.2 - 36.0) and admission creatinine >1.5 mg/dL (OR 16.9, 95% CI 10.4 - 27.3) were predictors of renal support. Moreover, Vallabhajosyula et al [9] reported that older age and greater comorbidity burden were associated with AKI and subsequent hemodialysis in patients with CS-AMI.

Contrary to expectations, NSTEMI presentation was associated with a higher risk of RRT compared to STEMI. This counterintuitive finding may reflect selection bias, where patients with NSTEMI developing CS represent a particularly high-risk subset with more comorbidities (diabetes and PAD) and chronic cardiac dysfunction. Alternatively, competing mortality risk may play a role, as patients with STEMI with severe shock may experience early death before developing AKI requiring RRT. This finding warrants further investigation in prospective studies and may indicate that patients with NSTEMI may benefit from more aggressive renal-protective strategies. This finding may also reflect differences in patient management strategies or in the underlying pathophysiology of AKI across subgroups. Patients with STEMI might benefit from more targeted revascularization strategies to mitigate renal injury. This is further supported by the CULPRIT-SHOCK trial, which demonstrated that, compared with immediate multivessel PCI, PCI of the culprit lesion was associated with a lower 30-day risk of death or severe renal failure requiring RRT [25].

A notable divergence exists between the variable importance rankings from the LASSO model and the significance findings from the cluster-robust logistic regression. This apparent contradiction reflects fundamental differences in the purpose and methodology of the 2 approaches rather than a true inconsistency. The LASSO model is optimized for prediction: it assigns coefficient weights based on the variable's contribution to discriminating between patients who will and will not require RRT across the entire dataset, without regard to statistical significance. The cluster-robust logistic regression, by contrast, is designed for inference: it estimates the independent association of each variable with the outcome while explicitly accounting for the nonindependence of observations within the same center. The application of cluster-robust (sandwich) standard errors inflates the variance of coefficient estimates to correct for within-center correlation, which reduces statistical power. The 2 analyses are complementary rather than contradictory, and the clinical utility of the nomogram is not diminished by the non-significance of individual predictors in the inferential model.

XSL•FO
RenderX

Clinical Implications

Our prediction model, presented as a user-friendly nomogram (Figure 4), is intended for use immediately post angiography, once the coronary anatomy has been defined. This represents a critical decision point in the management of CS-AMI, where clinicians must rapidly decide on the aggressiveness of hemodynamic support, the initiation of renal-protective strategies, and the need for early nephrology consultation. This early identification, aided by this tool, can trigger a cascade of preventative measures, including the avoidance of nephrotoxic agents, careful fluid management, and early involvement of a multidisciplinary team, including nephrologists, which has been shown to improve outcomes and mortality [10,11]. Finally, these data align with the growing evidence that early RRT initiation improves outcomes in selected patients.

Limitations

Several limitations warrant consideration. First, our complete-case analysis excluded patients with missing data, which may introduce selection bias. Second, the Cockcroft-Gault equation was used to estimate creatinine clearance; this formula assumes a steady-state serum creatinine level, which is rarely the case in the acute, dynamic setting of CS and may therefore underestimate the true degree of renal impairment. Third, the inclusion of the number of diseased vessels, an angiographic variable, means the model is intended for postcatheterization risk assessment and is not applicable for precatheterization risk stratification in the emergency department. Fourth, we did not perform a competing risks analysis to account for in-hospital mortality as a competing event for RRT because our registry lacks precise timing data for both events. Fifth, although the model was validated on an internal testing set and via bootstrapping, external validation in an independent cohort is needed before widespread clinical adoption. Sixth, a further limitation concerns the treatment of SCAI shock stage in the predictive model. The LASSO model retains a single continuous coefficient for SCAI stage, and the nomogram plots the stages on a single ordinal axis. This approach assumes equidistant and linear increments in risk between consecutive stages, which may not accurately reflect the disproportionate increase in RRT risk. However, the SCAI stage was included in the model as a single variable to simplify the nomogram for clinical use and to account for the small number of events in some SCAI stages. Finally, the wide confidence intervals observed for the SCAI stage comparisons are attributable to sparse outcome data in the reference category (stage B).

Conclusions

The study highlights significant differences in baseline characteristics and clinical outcomes between patients with CS who required renal support and those who did not. Key findings indicate that patients who needed renal support were older, had a greater prevalence of comorbidities, and presented with more severe clinical conditions, including higher rates of anterior STEMI and RV dysfunction. The analysis demonstrated that patients requiring renal replacement therapy had markedly higher hospital mortality rates and longer hospital stays, underscoring the gravity of their condition. Furthermore, the study identified critical predictors for RRT, including baseline creatinine clearance and advanced SCAI stage.

The development and internal validation of a 15-variable model allows accurate prediction of the need for RRT in patients with CS-AMI using only baseline data. Overall, these findings emphasize the importance of early identification and management of high-risk patients with CS to improve outcomes and tailor interventions more effectively. Further research may enhance the predictive capabilities and clinical applicability of these findings. Moreover, multidisciplinary management of CS, including nephrologists, is warranted.

Acknowledgments

The authors would like to thank all the nurses and cardiac catheterization technologists who helped in collecting the data from all the participating centers. The authors declare that generative artificial intelligence (AI) was used to improve the grammar, language, and structure of the manuscript.

Funding

The authors have no funding sources to disclose.

Data Availability

The data used in this study are available from the corresponding author upon justifiable requests and after institutional approval. As per institutions' protocols, data sharing requires a data sharing agreement.

Authors' Contributions

Study design, development, data collection: AD, JML, AA, AE, SS, HMA, MA, WA, PP, NY, AL, MAQ, WA, RR, MAK, AA, SAM, TH, JA, FE, MAA, MAJ, AE, BA, AMA, FOMN, HAN, MNA, AA, SAO, AAA, OK, AJ, YE, IAMA, MB, ASH, ASMAR, AFH, AA, MMNA, MAG, HR, AB, BA, AA, LA, SA, TSC, HK, SH, AAG, KA, RWA, AMT, AMS, AA, and AAA
Initial drafting of the manuscript: AD, JML, SS, FE, AA, SAO, AAA, OK, and AE
Analysis and interpretation of data: AD, AE, PP, NY, AA, AL, and AAA
Manuscript revision and final editing: AD, AE, WA, TH, JA, MAJ, PP, NY, AA, AL, and AAA
Statistician for this study: AAA

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables identified by Least Absolute Shrinkage and Selection Operator to be associated with renal replacement therapy.
[DOCX File, 15 KB - cardio_v10i1e79678_app1.docx]

Multimedia Appendix 2
Calibration plots before (left, blue) and after (right, green) Platt scaling. Each point represents a decile of predicted risk. Dashed line: perfect calibration. Platt scaling improved agreement between predicted and observed renal replacement therapy probabilities on the testing set (n=287).
[PNG File, 3731 KB - cardio_v10i1e79678_app2.png]

Multimedia Appendix 3
Bootstrap validation results (1000 iterations) showing the distribution of area under the receiver operating characteristic curve values from the training set (n=1072). Mean bootstrap area under the receiver operating characteristic curve 0.745 (red dashed line); 95% CI 0.730-0.756 (green dotted lines). The narrow CI demonstrates stable model performance.
[PNG File, 353 KB - cardio_v10i1e79678_app3.png]

References

1. Samsky MD, Morrow DA, Proudfoot AG, Hochman JS, Thiele H, Rao SV. Cardiogenic shock after acute myocardial infarction: a review. JAMA 2021 Nov 9;326(18):1840-1850. [doi: 10.1001/jama.2021.18323] [Medline: 34751704]
2. Vallabhajosyula S, Jentzer JC, Prasad A, et al. Epidemiology of cardiogenic shock and cardiac arrest complicating non-ST-segment elevation myocardial infarction: 18-year US study. ESC Heart Fail 2021 Jun;8(3):2259-2269. [doi: 10.1002/ehf2.13321] [Medline: 33837667]
3. Vallabhajosyula S, Bhopalwala HM, Sundaragiri PR, et al. Cardiogenic shock complicating non-ST-segment elevation myocardial infarction: an 18-year study. Am Heart J 2022 Feb;244:54-65. [doi: 10.1016/j.ahj.2021.11.002] [Medline: 34774802]
4. Adegbala O, Inampudi C, Adejumo A, et al. Characteristics and outcomes of patients with cardiogenic shock utilizing hemodialysis for acute kidney injury. Am J Cardiol 2019 Jun 1;123(11):1816-1821. [doi: 10.1016/j.amjcard.2019.02.038] [Medline: 30967283]
5. Ghionzoli N, Sciaccaluga C, Mandoli GE, et al. Cardiogenic shock and acute kidney injury: the rule rather than the exception. Heart Fail Rev 2021 May;26(3):487-496. [doi: 10.1007/s10741-020-10034-0] [Medline: 33006038]
6. Reinstadler SJ, Kronbichler A, Reindl M, et al. Acute kidney injury is associated with microvascular myocardial damage following myocardial infarction. Kidney Int 2017 Sep;92(3):743-750. [doi: 10.1016/j.kint.2017.02.016] [Medline: 28412022]
7. Lauridsen MD, Gammelager H, Schmidt M, et al. Acute kidney injury treated with renal replacement therapy and 5-year mortality after myocardial infarction-related cardiogenic shock: a nationwide population-based cohort study. Crit Care 2015 Dec 30;19:452. [doi: 10.1186/s13054-015-1170-8] [Medline: 26715162]
8. Marenzi G, Assanelli E, Campodonico J, et al. Acute kidney injury in ST-segment elevation acute myocardial infarction complicated by cardiogenic shock at admission. Crit Care Med 2010 Feb;38(2):438-444. [doi: 10.1097/CCM.0b013e3181b9eb3b] [Medline: 19789449]
9. Vallabhajosyula S, Dunlay SM, Barsness GW, et al. Temporal trends, predictors, and outcomes of acute kidney injury and hemodialysis use in acute myocardial infarction-related cardiogenic shock. PLoS One 2019;14(9):e0222894. [doi: 10.1371/journal.pone.0222894] [Medline: 31532793]
10. Zarbock A, Kellum JA, Schmidt C, et al. Effect of early vs delayed initiation of renal replacement therapy on mortality in critically Ill patients with acute kidney injury: the ELAIN randomized clinical trial. JAMA 2016;315(20):2190-2199. [doi: 10.1001/jama.2016.5828] [Medline: 27209269]
11. Zou H, Hong Q, Xu G. Early versus late initiation of renal replacement therapy impacts mortality in patients with acute kidney injury post cardiac surgery: a meta-analysis. Crit Care 2017 Jun 17;21(1):150. [doi: 10.1186/s13054-017-1707-0] [Medline: 28623953]
12. Patsalis N, Kreutz J, Chatzis G, et al. Early risk predictors of acute kidney injury and short-term survival during impella support in cardiogenic shock. Sci Rep 2024 Jul 30;14(1):17484. [doi: 10.1038/s41598-024-68376-w] [Medline: 39080441]
13. Daoulah A, Alshehri M, Panduranga P, et al. Clinical outcomes of patients with cardiogenic shock complicating acute myocardial infarction: the Gulf-Cardiogenic Shock registry. Shock 2024 Oct 1;62(4):512-521. [doi: 10.1097/SHK.0000000000002433] [Medline: 39158570]
14. Waksman R, Pahuja M, van Diepen S, et al. Standardized definitions for cardiogenic shock research and mechanical circulatory support devices: scientific expert panel from the Shock Academic Research Consortium (SHARC). Circulation 2023 Oct 3;148(14):1113-1126. [doi: 10.1161/CIRCULATIONAHA.123.064527] [Medline: 37782695]

XSL•FO
RenderX

15. Mehran R, Rao SV, Bhatt DL, et al. Standardized bleeding definitions for cardiovascular clinical trials. Circulation 2011 Jun 14;123(23):2736-2747. [doi: 10.1161/CIRCULATIONAHA.110.009449]
16. Daoulah A, Seraj S, Elmahrouk A, et al. Right ventricular dysfunction as a mortality determinant for patients with cardiogenic shock induced by acute myocardial infarction. Shock 2025 Jun 1;63(6):885-892. [doi: 10.1097/SHK.0000000000002583] [Medline: 40101947]
17. Sterling LH, Fernando SM, Talarico R, et al. Long-term outcomes of cardiogenic shock complicating myocardial infarction. J Am Coll Cardiol 2023 Sep 5;82(10):985-995. [doi: 10.1016/j.jacc.2023.06.026] [Medline: 37648357]
18. Jeppesen KK, Rasmussen SB, Kjaergaard J, et al. Acute kidney injury after out-of-hospital cardiac arrest. Crit Care 2024 May 18;28(1):169. [doi: 10.1186/s13054-024-04936-w] [Medline: 38762578]
19. Vallabhajosyula S, Ya'Qoub L, Kumar V, et al. Contemporary national outcomes of acute myocardial infarction-cardiogenic shock in patients with prior chronic kidney disease and end-stage renal disease. J Clin Med 2020 Nov 18;9(11):3702. [doi: 10.3390/jcm9113702] [Medline: 33218121]
20. Jain P, Thayer KL, Abraham J, et al. Right ventricular dysfunction is common and identifies patients at risk of dying in cardiogenic shock. J Card Fail 2021 Oct;27(10):1061-1072. [doi: 10.1016/j.cardfail.2021.07.013] [Medline: 34625126]
21. Bernhardt AM, Copeland H, Deswal A, et al. The International Society for Heart and Lung Transplantation/Heart Failure Society of America Guideline on Acute Mechanical Circulatory Support. J Heart Lung Transplant 2023 Apr;42(4):e1-e64. [doi: 10.1016/j.healun.2022.10.028] [Medline: 36805198]
22. Marenzi G, Cosentino N, Marinetti A, et al. Renal replacement therapy in patients with acute myocardial infarction: rate of use, clinical predictors and relationship with in-hospital mortality. Int J Cardiol 2017 Mar 1;230:255-261. [doi: 10.1016/j.ijcard.2016.12.130] [Medline: 28043673]
23. Coskun U, Orta Kilickesmez K, Abaci O, et al. The relationship between chronic kidney disease and SYNTAX score. Angiol Open Access 2011 Aug;62(6):504-508. [doi: 10.1177/0003319711398864] [Medline: 21422054]
24. Freund A, Jobs A, Lurz P, et al. Frequency and impact of bleeding on outcome in patients with cardiogenic shock. JACC Cardiovasc Interv 2020 May 25;13(10):1182-1193. [doi: 10.1016/j.jcin.2020.02.042] [Medline: 32438988]
25. Thiele H, Akin I, Sandri M, et al. PCI strategies in patients with acute myocardial infarction and cardiogenic shock. N Engl J Med 2017 Dec 21;377(25):2419-2432. [doi: 10.1056/NEJMoa1710261] [Medline: 29083953]

Abbreviations

AKI: acute kidney injury
AMI: acute myocardial infarction
AUC: area under the receiver operating characteristic curve
BARC: Bleeding Academic Research Consortium
CABG: coronary artery bypass grafting
CS: cardiogenic shock
CS-AMI: cardiogenic shock secondary to acute myocardial infarction
CVA: cerebrovascular accident
ELAIN: Early versus Late Initiation of RRT in Critically Ill Patients with Acute Kidney Injury
Gulf-CS: Gulf Cardiogenic Shock
IRB: Institutional Review Board
LASSO: Least Absolute Shrinkage and Selection Operator
MCS: mechanical circulatory support
NSTEMI: non–ST-elevation myocardial infarction
OR: odds ratio
PAD: peripheral arterial disease
PCI: percutaneous coronary intervention
RRT: renal replacement therapy
RV: right ventricular
SCAI: Society for Cardiovascular Angiography and Interventions
SCAI-CSWG: Society for Cardiovascular Angiography and Interventions Cardiogenic Shock Working Group
STEMI: ST-elevation myocardial infarction
VIF: variance inflation factor

XSL•FO
RenderX

Edited by A Coristine; submitted 25.Jun.2025; peer-reviewed by M Alrefai; revised version received 06.Apr.2026; accepted 07.Apr.2026; published 20.May.2026.

Please cite as:
Daoulah A, M Ladna J, Arabi A, Elmahrouk A, Seraj S, M Aloui H, Alshehri M, Almahmeed W, Panduranga P, Yousif N, Lotfi A, Qutub M, Alharbi W, Rajan R, Abdirahman Kahin M, Alenezi A, Al Maashani S, Hassan T, Alswuaidi J, Eltaieb F, Abdulhadi Aldossari M, Al Jarallah M, Elganady A, Alzahrani B, Alqahtani A, Omar M Al Nasser F, A Noor H, Alama MN, Aldossari A, Al Obaikan S, Ali Almarghany A, Kanbr O, Elmahrouk Y, A M Abdulhabeeb I, Balghith M, S Hersi A, Antony H, Fathey Hussien A, Almoghairi A, Alama MMA, Ajaz Ghani M, Rahbi H, Basardah A, Alshehri B, Alama A, AlObaid L, Abualnaja S, Shahzad Chachar T, Khan H, Hashmani S, A Ghonim A, Almerri K, Alsofayan R, M Tawengi A, M Shawky A, Alshehri A, Arafat A, Jamjoom A
Postangiography Prediction of Renal Replacement Therapy in Acute Myocardial Infarction–Related Cardiogenic Shock: Least Absolute Shrinkage and Selection Operator Nomogram Development and Validation
JMIR Cardio 2026;10:e79678
URL: https://cardio.jmir.org/2026/1/e79678
doi:10.2196/79678

© Amin Daoulah, Julia M Ladna, Abdulrahman Arabi, Ahmed Elmahrouk, Shaber Seraj, Hatem M Aloui, Mohammed Alshehri, Wael Almahmeed, Prashanth Panduranga, Nooraldaem Yousif, Amir Lotfi, Mohammed Qutub, Waleed Alharbi, Rajesh Rajan, Mokhtar Abdirahman Kahin, Abdullah Alenezi, Said Al Maashani, Taher Hassan, Jassim Alswuaidi, Fakhreldein Eltaieb, Mubarak abdulhadi Aldossari, Mohammed Al Jarallah, Abdelmaksoud Elganady, Badr Alzahrani, Abdulrahman Alqahtani, Faisal Omar M Al Nasser, Husam A Noor, Mohamed N Alama, Alaa Aldossari, Sultan Al Obaikan, Alsayed Ali Almarghany, Omar Kanbr, Youssef Elmahrouk, Ibrahim A M Abdulhabeeb, Mohammed Balghith, Ahmad S Hersi, Harvey Antony, Adnan Fathey Hussien, Abdulrahman Almoghairi, Mohamed Mohammednabil A Alama, Mohamed Ajaz Ghani, Hazim Rahbi, Ayman Basardah, Bandar Alshehri, Abdulhadi Alama, Laura AlObaid, Seraj Abualnaja, Tarique Shahzad Chachar, Hassan Khan, Shahrukh Hashmani, Ahmed A Ghonim, Khalid Almerri, Razan Alsofayan, Abdelaziz M Tawengi, Abeer M Shawky, Ali Alshehri, Amr Arafat, Ahmed Jamjoom. Originally published in JMIR Cardio (https://cardio.jmir.org), 20.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Operationalizing AI-Enabled Cardiovascular Biomarkers: A Clinician-Centered Framework for Validation, Governance, and Workflow Integration

Dabeluchi Chiedozie Ngwu, MD

Department of Surgery, King Abdullah Hospital, Al Nakhil, Bisha, Asir, Saudi Arabia

Corresponding Author:
Dabeluchi Chiedozie Ngwu, MD
Department of Surgery, King Abdullah Hospital, Al Nakhil, Bisha, Asir, Saudi Arabia

Abstract

Cardiovascular biomarkers are increasingly extracted or interpreted using artificial intelligence (AI) applied to electrocardiograms, imaging, laboratory measurements, electronic health records, wearable devices, and longitudinal data. Predictive performance alone, however, does not establish that a measurement is valid, that a model is transportable and calibrated, or that acting on its output improves care. Existing resources provide essential but complementary foundations: the US Food and Drug Administration–National Institutes of Health (FDA-NIH) BEST (Biomarkers, EndpointS, and other Tools) resource standardizes biomarker terminology; the V3 and V3+ frameworks address verification, analytical validation, clinical validation, and usability of digitally measured signals; prediction-model and trustworthy-AI guidance addresses reporting, risk of bias, early clinical evaluation, and deployability; and regulatory qualification pathways evaluate evidence within a defined context of use. A remaining practical challenge is translating these complementary requirements into accountable clinical action within cardiovascular workflows. This viewpoint proposes a clinician-centered operational framework organized around validation, governance, and workflow integration. For each pillar, it identifies accountable actors, required steps, documented outputs, and escalation or stop rules. Validation establishes whether the input measurement and model are fit for the intended population and decision. Governance assigns institutional authorization, clinical ownership, version control, monitoring, and authority to restrict, pause, or withdraw the intervention. Workflow integration specifies who receives the output, what confirmatory action follows, how disagreement is handled, and how decisions are documented. A worked example of an AI-enabled electrocardiographic screening output for possible left ventricular systolic dysfunction illustrates the pathway from local evaluation to echocardiographic confirmation and lifecycle monitoring. The framework does not replace established validation or regulatory standards; it operationalizes them as a clinician-centered and institutionally accountable pathway from validated signal to governed, patient-facing action.

(JMIR Cardio 2026;10:e103401) doi:10.2196/103401

KEYWORDS

artificial intelligence; cardiovascular biomarkers; digital biomarker; clinical validation; clinical decision support systems; workflow integration; governance

Introduction

Cardiovascular biomarkers influence consequential decisions, including diagnosis, triage, preventive treatment, referral, device therapy, and longitudinal monitoring. Artificial intelligence (AI) is expanding this landscape by extracting or interpreting signals from electrocardiograms (ECGs), imaging, laboratory measurements, electronic health records (EHRs), wearable devices, omics platforms, and longitudinal clinical data. These methods may reveal patterns that conventional thresholds or models do not capture and support earlier recognition or more individualized assessment [1-12]. Their clinical value, however, depends on more than performance in a development dataset.

Terminology is important because several related objects are often conflated. Following the US Food and Drug Administration–National Institutes of Health (FDA-NIH) BEST (Biomarkers, EndpointS, and other Tools) resource, a biomarker is a defined characteristic measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention [13]. A digital biomarker is a biomarker measured or derived using digital technology [14,15]. In this viewpoint, an AI-enabled cardiovascular biomarker is a cardiovascular biomarker whose extraction, quantification, combination, or interpretation materially depends on an AI method. Digital biomarkers and AI-enabled biomarkers overlap, but neither wholly contains the other: a digitally captured heart-rate signal may require no AI, whereas AI may interpret conventional laboratory or imaging measurements. A prediction or risk score is a model-generated estimate and is not automatically a biomarker. A complete clinical decision-support intervention comprises the data input, model or interpretive

XSL•FO
RenderX

process, output, interface, intended user, response pathway, and oversight arrangements.

These distinctions matter because an AI-enabled biomarker output may be associated with disease without being suitable for clinical action. A model may discriminate well yet be poorly calibrated, lose performance across settings or patient groups, or trigger testing and treatment whose benefits and harms have not been established. Reliability concerns the consistency or reproducibility of a measurement, whereas validity concerns whether it accurately captures the intended characteristic. External validation tests performance in data separated from development by place, time, population, or setting. Calibration concerns agreement between predicted probabilities and observed event frequencies in the intended population, particularly around thresholds used for action. Clinical utility asks whether using the output improves decisions or patient-relevant outcomes enough to justify its consequences [2-4,16-23]. Predictive performance in a development dataset is therefore only one component of validation and does not establish measurement validity, transportability, calibration, or clinical utility in the intended setting.

Existing frameworks provide complementary foundations. BEST standardizes biomarker terminology and categories [13]. V3 addresses verification, analytical validation, and clinical validation for biometric monitoring technologies, while V3+ extends this foundation to usability validation for sensor-based digital health technologies [14,15]. Cardiovascular guidance addresses the predictive and clinical utility of novel biomarkers and models [2-4]. TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence) and PROBAST+AI (Prediction Model Risk of Bias Assessment Tool + Artificial Intelligence) address reporting, risk of bias, and applicability; DECIDE-AI (Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence) addresses early-stage clinical evaluation; and FUTURE-AI (Fairness, Universality, Traceability, Usability, Robustness, and Explainability in Artificial Intelligence) addresses trustworthy and deployable AI [23-26]. FDA and European Medicines Agency qualification pathways also anchor evidence to a defined context of use in medical product development [27,28]. V3+ focuses on usability, whereas the workflow-integration pillar considered here addresses the broader clinical pathway through which an output is received, confirmed, acted upon, documented, and monitored.

This viewpoint does not propose a new biomarker taxonomy or replace established validation standards. It integrates these complementary requirements into a clinician-centered implementation pathway that links fit-for-purpose evidence to named clinical and institutional responsibilities, predefined clinical responses, and lifecycle monitoring. The framework is organized around 3 interconnected pillars: validation, governance, and workflow integration. A visual abstract of the proposed framework is presented in Multimedia Appendix 1.

From Biomarker Signal to Clinical Intervention

Biomarker translation can fail before AI is introduced. The underlying assay, sensor, image acquisition process, or data stream may be unstable; the development cohort may be narrow; the outcome may be poorly defined; or the proposed use may not correspond to a meaningful clinical decision. Sample handling, device characteristics, preprocessing, timing, biological variability, kidney function, comorbidity, medication use, and missing data can all alter the behavior of a cardiovascular signal. If the input is unreliable or invalid, a more sophisticated model cannot repair the evidentiary foundation [2,3,16-19].

Transportability presents a second challenge. Cardiovascular populations and workflows vary across hospitals, laboratories, ECG systems, imaging platforms, EHRs, referral practices, and treatment pathways. Average performance may conceal important failures across age, sex, comorbidity, ancestry or ethnicity (when these data are appropriately collected), kidney function, device access, and care setting. Because local prevalence also changes positive and negative predictive values, a threshold developed in a tertiary referral cohort may generate an unacceptable false-positive burden in primary care. External and local validation must therefore examine calibration, subgroup performance, and the consequences of the proposed threshold, not discrimination alone [23,24,29-31].

Clinical utility remains distinct from statistical performance. A reproducible and externally validated output may add little beyond existing assessment, identify patients for whom no beneficial response is available, or generate unnecessary testing, anxiety, treatment, and cost. Incremental performance, decision-curve analysis, pragmatic evaluation, and patient-relevant outcomes can help determine whether acting on the output is preferable to current care [4,17,20,22,32]. The relevant unit of evaluation is therefore the complete clinical decision-support intervention: who receives the output, what action follows, and what happens to the patient.

AI can nevertheless add meaningful capabilities by combining multimodal signals, identifying nonlinear relationships, supporting phenotype discovery, and enabling longitudinal interpretation. It may also support lifecycle monitoring by detecting calibration drift, changes in data quality, and subgroup deterioration. These capabilities become clinically meaningful only when the output is tied to a defined intended use, accountable ownership, and an evaluable response pathway. Otherwise, AI may scale weak evidence or ambiguous action more rapidly than conventional biomarker development [5,8-12,21,25,26,32-34].

A Clinician-Centered Operational Framework

Overview

The 3 pillars are connected, rather than parallel, checklists. Validation determines what the output can legitimately support.

XSL•FO
RenderX

Governance determines who may authorize, monitor, modify, pause, or withdraw it. Workflow integration determines what happens when the output enters clinical care. Each pillar is therefore defined by accountable actors, required steps, documented outputs, and escalation or stop rules.

Validation

Accountable actors include the developer or vendor, a local clinical lead, an expert in the relevant measurement domain, and a biostatistics or data-science team. Their first task is to define the intended use: the target population, setting, decision, timing, comparator, and actions that may follow. Verification and analytical validation should establish that the acquisition system, assay, device, preprocessing pipeline, and data transfer perform as intended. Clinical validation should establish the relationship between the measured characteristic or model output and the relevant clinical state or outcome [13-15].

Where an AI model produces a probability, classification, or risk score, evaluation should include separation from development data, discrimination, calibration, clinically relevant subgroup performance, robustness to local devices and data pipelines, and positive or negative predictive value in the intended population. Thresholds should be evaluated in relation to downstream action rather than selected for statistical convenience. Clinical utility should determine whether the output adds value beyond current assessment and whether expected benefits justify confirmatory testing, referral, treatment, or monitoring [4,20,22-24,26,27,32]. When important subgroup differences are identified, evaluation should move beyond reporting them to determining whether data revision, threshold adjustment, workflow safeguards, restricted use, or nondeployment is required.

The resulting local evidence report should specify the approved population, exclusions, model version, data requirements, threshold, uncertainty, expected downstream burden, and residual limitations. Deployment should not proceed when the input is insufficiently reliable or valid, performance is not transportable or calibrated for the intended setting, important subgroup risks remain unresolved, or no justifiable clinical response follows.

Governance

Governance requires both a named clinical service owner and a multidisciplinary institutional oversight function. The exact structure may vary, but it should include clinical, informatics, data-science, quality and safety, privacy, and legal or risk-management expertise, with patient or public perspectives incorporated when warranted by the intended use and patient-facing consequences. Vendors may provide technical documentation and model updates, but the institution retains responsibility for local authorization and oversight and should maintain an independent record of the deployed version.

Before deployment, the oversight body should approve the intended use, eligible population, prohibited uses, data dependencies, clinical owner, monitoring metrics, review cadence, incident-reporting route, and procedure for material updates. Version changes, altered input pipelines, new devices, population shifts, or workflow redesign may invalidate prior assumptions and should trigger proportionate reassessment. Prespecified monitoring criteria may include material calibration deterioration, unacceptable subgroup disparities, increased false-positive burden, clinically important missed cases, harmful downstream action, or substantial changes in data quality or software behavior [25,26,32-37].

Governance should produce formal authorization, a named accountable owner, a version and change log, a monitoring protocol, and procedures for recalibration, restricted use, temporary suspension, and withdrawal. It is effective only when someone has both the information and authority to act when performance or clinical consequences become unacceptable.

Workflow Integration

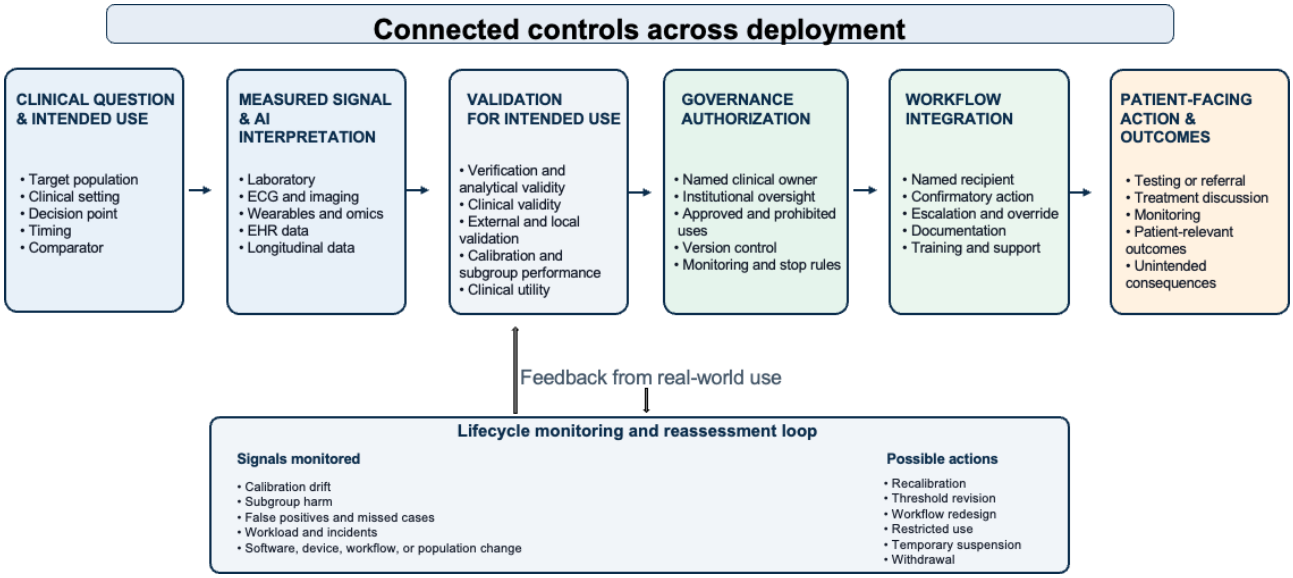
Workflow integration is owned by the clinical pathway rather than the algorithm. The clinical service lead, intended users, informatics team, and relevant nursing or allied professionals should determine who receives the output, when and where it appears, what confirmatory action is appropriate, how urgency is communicated, and how uncertainty or disagreement is documented. The output should be designed around a decision point rather than added as an unprioritized dashboard item or interruptive alert.

A defined response pathway specifies the recipient; the threshold or conditions that trigger review; permissible actions; any required confirmation before treatment; escalation and de-escalation rules; when an override rationale is required; and how responsibility transfers across teams. Documentation should be proportionate to clinical risk. When an AI output conflicts with clinical judgment, the pathway should preserve clinician authority while making high-risk disagreements visible for review. Patient communication should be specified when the result leads to testing, referral, or treatment discussion.

The pathway should be written, tested, and embedded in the relevant clinical system. Implementation readiness also requires role-specific training so that intended users understand the output, its limitations, the required confirmatory actions, and the circumstances in which escalation or override is appropriate. Deployment should be delayed when the recipient is unclear, confirmatory capacity is unavailable, the alert cannot be acted on within the required time, the output encourages unsupported treatment, or responsibility is fragmented across teams. Workflow simulation, silent or shadow-mode testing, and early prospective evaluation can identify these failures before broad implementation [9-12,26,32]. Figure 1 summarizes the clinician-centered operational pathway from intended use and signal measurement through validation, governance authorization, workflow integration, patient-facing action, and lifecycle reassessment.

XSL•FO
RenderX

Figure 1. Clinician-centered operational pathway for AI-enabled cardiovascular biomarkers. AI: artificial intelligence; ECG: electrocardiogram; EHR: electronic health record.



The pathway begins with a defined clinical question and intended use, followed by measurement and AI-enabled interpretation, validation for the intended population and decision, governance authorization, workflow integration, and patient-facing action. Validation, governance, and workflow integration operate as connected controls across deployment. Postdeployment monitoring evaluates calibration drift, subgroup harm, false-positive and missed-case patterns, workload, incidents, and changes in software, devices, workflows, or populations. Findings may trigger recalibration, threshold revision, workflow redesign, restricted use, temporary suspension, or withdrawal. The pathway, therefore, does not end at model validation; authorization, clinical response, and lifecycle reassessment determine whether an output becomes a safe clinical decision-support intervention.

Applying the Framework: AI-ECG Screening for Left Ventricular Systolic Dysfunction

Consider an AI-enabled ECG model intended to identify patients with possible left ventricular systolic dysfunction who may benefit from echocardiography. Development studies have shown that AI can detect a signal associated with low ejection fraction from a standard ECG, and pragmatic evaluation has shown that presenting this output in routine primary care can increase detection of low ejection fraction [38,39]. Its intended use must remain narrow: the output is a screening prompt for confirmatory assessment, not a diagnosis of heart failure or an independent indication for treatment.

Validation would require confirmation that the locked model is compatible with local ECG acquisition, preprocessing, and data-transfer systems, followed by testing in a local or otherwise suitably representative population independent of model development. Evaluation should include discrimination, calibration, positive predictive value, subgroup performance, and the expected number of echocardiograms generated per confirmed case. A threshold acceptable in a referral population may produce excessive false positives or overwhelm echocardiography capacity in lower-prevalence primary care.

Governance would require institutional authorization, a named clinical owner, and documented responsibility for version control and monitoring. Relevant measures would include false-positive and false-negative patterns, subgroup performance, echocardiography yield, time to confirmation, alert burden, and inappropriate downstream action. Material performance deterioration or a change in the software or data pipeline should trigger reassessment and, when necessary, restricted use or suspension.

Within the workflow, the output would reach a named ordering clinician or designated results-management team with a clear recommendation for clinical review and, where appropriate, confirmatory echocardiography. The pathway would specify urgency, referral options, documentation, and how to respond when recent imaging or an alternative explanation is already available. No diagnosis of heart failure or disease-specific treatment should follow solely from the AI-ECG output. The example illustrates how a technically valid prediction becomes part of a safer clinical decision-support intervention only when evidence, ownership, confirmation, and monitoring are connected.

Operational Requirements

Table 1 translates the 3 pillars and the cross-cutting lifecycle-monitoring function into minimum operational requirements. The requirements are role-based rather than tied to a single committee structure because institutions differ in size and resources. Across settings, explicit ownership, documented evidence, a tested response pathway, and authority to restrict or stop use should remain nonnegotiable.

XSL•FO
RenderX

Table . Minimum operational requirements for validation, governance, workflow integration, and lifecycle monitoring of AI-enabled cardiovascular biomarkers.

Domain	Accountable actors	Minimum operational requirements	Required outputs and action threshold
Validation	Developer or vendor; local clinical lead; measurement-domain expert; biostatistics or data-science team	Define intended use and population; verify acquisition and preprocessing; conduct analytical and clinical validation; assess transportability, calibration, subgroup performance, and clinical utility	Local evidence report specifying population, exclusions, model version, decision threshold, uncertainty, and limitations. Do not deploy when validity, calibration, transportability, subgroup safety, or actionable utility is inadequate.
Governance	Named clinical service owner; multidisciplinary institutional oversight body; quality, safety, and informatics representatives	Authorize intended and prohibited uses; document data dependencies and version; assign monitoring; review incidents and material updates; define restriction, pause, and withdrawal procedures	Authorization record, accountable owner, change log, and monitoring plan. Reassess, restrict, pause, or withdraw use after material performance deterioration, inequity, harmful consequences, or major system change.
Workflow integration	Clinical pathway owner; intended clinicians; informatics; nursing and allied professionals as relevant	Identify recipient, timing, interface, confirmatory action, urgency, escalation, override, documentation, patient communication, and training requirements	Written and tested response pathway, workflow map, and training materials. Delay deployment when responsibility is unclear, action capacity is unavailable, alert burden is unsafe, or unsupported treatment may result.
Lifecycle monitoring (cross-cutting)[a]	Clinical owner; governance body; analytics or quality team; vendor as applicable	Track calibration, yield, data quality, subgroup performance, missed cases, false positives, downstream actions, workload, incidents, and version changes	Monitoring report and corrective-action record. Trigger recalibration, threshold revision, workflow redesign, restricted use, suspension, or withdrawal according to prespecified risk-based criteria.

[a]Lifecycle monitoring is cross-cutting and supports all 3 pillars.

Implications for Evaluation and Reporting

Investigators, peer reviewers, health systems, and regulators should evaluate AI-enabled cardiovascular biomarker studies beyond discrimination alone. Reports should clarify whether the object under evaluation is a biomarker, digital biomarker, prediction model, or complete clinical decision-support intervention; define the intended use and target population; describe measurement verification and validity; report external validation, calibration, and clinically relevant subgroup performance; and explain the comparator and downstream consequences. When deployment is proposed, reports should also identify the accountable owner, response pathway, prohibited uses, monitoring plan, and criteria for recalibration, suspension, or withdrawal.

The evidentiary question, therefore, shifts from "Does the model predict?" to "Are the measured signal, model, institution, and response pathway jointly fit for the intended use?" A valid measurement and a well-performing model do not by themselves establish the clinical value or safety of deployment, and health systems should not infer either from regulatory status, technical performance, or publication. Prospective evaluation should examine patient-relevant outcomes, unintended consequences, clinician workload, equity, and the behavior of the complete clinical decision-support intervention over time [25,26,32].

Conclusion

AI can expand the information obtainable from cardiovascular biomarkers and related clinical data, but technical performance alone does not establish that an input is valid, an output is transportable and calibrated, or that acting on it improves care.

The proposed framework links established evidentiary standards to clinical implementation by connecting validation, accountable governance, defined response pathways, and lifecycle monitoring. The worked AI-ECG example shows why these functions must be designed together. Clinical trust should therefore be earned at the level of the complete clinical decision-support intervention: a validated signal, a fit-for-purpose model, accountable clinical and institutional ownership, a defined response pathway, and continuing evidence that the intervention remains safe and useful in practice.

Acknowledgments

The author thanks colleagues and external reviewers who provided informal feedback during manuscript development. The author also acknowledges the faculty and course contributors of the MIT Artificial Intelligence in Health Care program and the Harvard

XSL•FO
RenderX

programs in health care transformation, whose teaching and discussions informed the author's thinking about clinical artificial intelligence, implementation, and organizational change.

Funding

No funding was received for this work.

Data Availability

No new data were generated or analyzed in this article. Data sharing is not applicable because this is a viewpoint based on the interpretation of published literature and the development of a conceptual framework.

Authors' Contributions

DCN conceived the manuscript, developed the clinician-centered framework, drafted the manuscript, revised the intellectual content, and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Visual abstract.
[PNG File, 118 KB - cardio_v10i1e103401_app1.png]

References

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001 Mar;69(3):89-95. [doi: 10.1067/mcp.2001.113989] [Medline: 11240971]
2. Morrow DA, de Lemos JA. Benchmarks for the assessment of novel cardiovascular biomarkers. Circulation 2007 Feb 27;115(8):949-952. [doi: 10.1161/CIRCULATIONAHA.106.683110] [Medline: 17325253]
3. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation 2009 May 5;119(17):2408-2416. [doi: 10.1161/CIRCULATIONAHA.109.192278] [Medline: 19364974]
4. Khan SS, Greenland P, Hayman LL, et al. Criteria to assess the predictive and clinical utility of novel models, biomarkers, and tools for risk of cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2026 Mar 17;153(11):e953-e970. [doi: 10.1161/CIR.0000000000001401] [Medline: 41669831]
5. Krittanawong C, Rogers AJ, Johnson KW, et al. Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management. Nat Rev Cardiol 2021 Feb;18(2):75-91. [doi: 10.1038/s41569-020-00445-9] [Medline: 33037325]
6. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 7;394(10201):861-867. [doi: 10.1016/S0140-6736(19)31721-0] [Medline: 31378392]
7. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
8. Chaturvedi A, Prabhakaran D. Transforming cardiovascular care with digital health: the past, progress, and promise. JACC Adv 2024 Sep;3(9):101183. [doi: 10.1016/j.jacadv.2024.101183] [Medline: 39220713]
9. Stremmel C, Breitschwerdt R. Digital transformation in the diagnostics and therapy of cardiovascular diseases: comprehensive literature review. JMIR Cardio 2023 Aug 30;7:e44983. [doi: 10.2196/44983] [Medline: 37647103]
10. Abedi A, Verma A, Jain D, et al. AI-driven real-time monitoring of cardiovascular conditions with wearable devices: scoping review. JMIR Mhealth Uhealth 2025 Nov 11;13:e73846. [doi: 10.2196/73846] [Medline: 41218203]
11. Bozyel S, Şimşek E, Koçyiğit Burunkaya D, et al. Artificial intelligence-based clinical decision support systems in cardiovascular diseases. Anatol J Cardiol 2024 Jan 7;28(2):74-86. [doi: 10.14744/AnatolJCardiol.2023.3685] [Medline: 38168009]
12. Cho Y, Yoon M, Kim J, et al. Artificial intelligence-based electrocardiographic biomarker for outcome prediction in patients with acute heart failure: prospective cohort study. J Med Internet Res 2024 Jul 3;26:e52139. [doi: 10.2196/52139] [Medline: 38959500]
13. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource: Food and Drug Administration (US); National Institutes of Health (US); 2016. URL: https://www.ncbi.nlm.nih.gov/books/NBK326791 [accessed 2026-07-14]

XSL•FO
RenderX

14. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
15. Bakker JP, Barge R, Centra J, et al. V3+ extends the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies. NPJ Digit Med 2025 Jan 24;8(1):51. [doi: 10.1038/s41746-024-01322-2] [Medline: 39856145]
16. Institute of Medicine (US) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease. In: Micheel CM, Ball JR, editors. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease: National Academies Press (US); 2010. [doi: 10.17226/12869] [Medline: 25032382]
17. Ioannidis JPA, Bossuyt PMM. Waste, leaks, and failures in the biomarker pipeline. Clin Chem 2017 May;63(5):963-972. [doi: 10.1373/clinchem.2016.254649] [Medline: 28270433]
18. Saleh S, George J, Kott KA, Meikle PJ, Figtree GA. The translation and commercialisation of biomarkers for cardiovascular disease-a review. Front Cardiovasc Med 2022;9:897106. [doi: 10.3389/fcvm.2022.897106] [Medline: 35722087]
19. Cunningham JW, Pfeffer MA. Clinical translation of cardiovascular biomarkers - a long and uncertain road. NEJM Evid 2022 Oct;1(10):EVIDe2200222. [doi: 10.1056/EVIDe2200222] [Medline: 38319843]
20. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3(1):18. [doi: 10.1186/s41512-019-0064-7] [Medline: 31592444]
21. Zehra B, Vinod N, BinEshaq S, et al. Multiomics approaches to cardiovascular disease: technological innovations and clinical translation. Am J Physiol Heart Circ Physiol 2026 Jun 1;330(6):H1752-H1784. [doi: 10.1152/ajpheart.00024.2026] [Medline: 42013055]
22. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008 Jan 30;27(2):157-172. [doi: 10.1002/sim.2929] [Medline: 17569110]
23. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ 2025 Mar 24;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
24. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
25. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ 2025 Feb 5;388:e081554. [doi: 10.1136/bmj-2024-081554] [Medline: 39909534]
26. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med 2022 May;28(5):924-933. [doi: 10.1038/s41591-022-01772-9] [Medline: 35585198]
27. Context of use. US Food and Drug Administration. 2026 Jan 21. URL: https://www.fda.gov/drugs/biomarker-qualification-program/context-use [accessed 2026-07-14]
28. Qualification of novel methodologies for medicine development. European Medicines Agency. URL: https://www.ema.europa.eu/en/human-regulatory-overview/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development [accessed 2026-07-14]
29. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
30. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872. [doi: 10.7326/M18-1990] [Medline: 30508424]
31. Yagi R, Goto S, Katsumata Y, MacRae CA, Deo RC. Importance of external validation and subgroup analysis of artificial intelligence in the detection of low ejection fraction from electrocardiograms. Eur Heart J Digit Health 2022 Dec;3(4):654-657. [doi: 10.1093/ehjdh/ztac065] [Medline: 36710903]
32. Jain SS, Goto S, Hall JL, et al. Pragmatic approaches to the evaluation and monitoring of artificial intelligence in health care: a science advisory from the American Heart Association. Circulation 2025 Dec 9;152(23):e433-e442. [doi: 10.1161/CIR.0000000000001400] [Medline: 41208719]
33. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
34. Davis SE, Lasko TA, Chen G, Matheny ME. Calibration drift among regression and machine learning models for hospital mortality. AMIA Annu Symp Proc 2017;2017:625-634. [Medline: 29854127]
35. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. EUR-Lex. 2025 Feb 11. URL: http://data.europa.eu/eli/reg/2025/327/oj [accessed 2026-05-31]
36. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). EUR-Lex. 2024 Jun 13. URL: http://data.europa.eu/eli/reg/2024/1689/oj [accessed 2026-05-31]

XSL•FO
RenderX

37. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. EUR-Lex. 2017 Apr 5. URL: http://data.europa.eu/eli/reg/2017/745/oj [accessed 2026-05-31]
38. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019 Jan;25(1):70-74. [doi: 10.1038/s41591-018-0240-2] [Medline: 30617318]
39. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med 2021 May;27(5):815-819. [doi: 10.1038/s41591-021-01335-4] [Medline: 33958795]

Abbreviations

AI: artificial intelligence
BEST: Biomarkers, EndpointS, and other Tools
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence
ECG: electrocardiogram
EHR: electronic health record
FDA: US Food and Drug Administration
FUTURE-AI: Fairness, Universality, Traceability, Usability, Robustness, and Explainability in Artificial Intelligence
NIH: National Institutes of Health
PROBAST+AI: Prediction Model Risk of Bias Assessment Tool + Artificial Intelligence
TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence

Edited by F Muehlensiepen; submitted 02.Jun.2026; peer-reviewed by AA Arafat, R Lambe; accepted 11.Aug.2026; published 28.Aug.2026.

Please cite as:
Ngwu DC
Operationalizing AI-Enabled Cardiovascular Biomarkers: A Clinician-Centered Framework for Validation, Governance, and Workflow Integration
JMIR Cardio 2026;10:e103401
URL: https://cardio.jmir.org/2026/1/e103401
doi:10.2196/103401

© Dabeluchi Chiedozie Ngwu. Originally published in JMIR Cardio (https://cardio.jmir.org), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Addressing the Challenges in Using Synthetic Data for Health Research: Application to Cardiology

Louise Baschet[1], MSc; Sebastien Marque[1], PhD; Stéphane Locret[2], MD; Jens Jenssen[3], MD; Vanessa Barbet[1], MSc; Patrick Jourdain[4], MSc, MBA, MD

[1]Horiana, 80 bis rue Paul Camelle, Bordeaux, France

[2]Research Department, Ramsay Santé, Paris, France

[3]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet and Capio St Görans Hospital, Stockholm, Sweden

[4]Medical Department, Ramsay Sante, Paris Saclay University, Paris, France

Corresponding Author:
Louise Baschet, MSc
Horiana, 80 bis rue Paul Camelle, Bordeaux, France

Abstract

Synthetic data offer significant potential for cardiology research by enabling data sharing, preserving privacy, and supporting machine learning model development. By generating artificial patient records that reflect real-world distributions, synthetic data can accelerate clinical research, improve model performance for rare cardiovascular conditions, and facilitate transnational collaborations that would otherwise be restricted by data-sharing barriers. Despite these advantages, the increasing use of synthetic data raises important ethical, regulatory, and methodological concerns that remain insufficiently addressed. Key challenges include assessing the validity and generalizability of synthetic datasets, understanding their limitations in representing complex and heterogeneous patient populations, and preventing the amplification of existing biases in cardiovascular care. Current regulatory frameworks, including the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), do not fully address emerging risks such as reidentification and data leakage, and there is no harmonized guidance to govern the use of synthetic data as stand-alone evidence for medical device evaluation or therapeutic research. In this viewpoint, we argue that responsible integration of synthetic data in cardiology requires, first, clear differentiation between synthetic data as a privacy-preserving distributional substitute and synthetic data as a counterfactual simulation tool, and, second, fit-for-purpose governance frameworks that pair rigorous utility and fidelity testing with explicit, adversary-aware privacy evaluation before synthetic cohorts are accepted as evidence in research or product evaluation. A prerequisite for that governance is conceptual clarity about what synthetic data are being used for. Synthetic data in health care serve 2 fundamentally distinct roles that carry entirely different validity requirements, failure modes, and regulatory implications, yet they are routinely conflated. The first role is as a privacy-preserving distributional substitute: the goal is statistical fidelity to the real data distribution, so that analyses of the synthetic dataset yield results equivalent to those of the original. The second role is as a tool for counterfactual simulation: the goal is to generate data that could not have been observed, such as rare conditions, hypothetical interventions, or extrapolations to new populations. These 2 roles are methodologically distinct. A dataset that accurately reflects real-world distributions may be inadequate for extrapolating findings to underrepresented subgroups. Conversely, a simulator optimized for novel scenario generation may systematically diverge from real-world distributions. This distinction informs every subsequent discussion of validity, bias, and regulation in this viewpoint and our proposed 4 concrete actions for the cardiology research community, including mandatory 3-layer (fidelity, utility, and privacy) validation, systematic subgroup reporting, explicit intended-use scoping, and domain-specific acceptability thresholds for synthetic data–based evidence.

(JMIR Cardio 2026;10:e92930) doi:10.2196/92930

KEYWORDS

health care; machine learning; synthetic data; virtual patients; cardiology; medical research

Introduction: Synthetic Data as an Emerging Tool in Cardiology

Used for many decades in other fields, synthetic data were adopted more recently in health care and have been shown to offer significant advancements. Specifically, synthetic data address challenges related to data privacy, progressively optimize care, and improve clinical research [1-3]. In oncology, models trained on synthetic data have been shown to predict survival more accurately than models trained on original data [4]. In addition, use of massive data can accelerate research, reducing costs and inclusion periods associated with clinical trials and data analysis. However, while the benefits are now becoming well known, the limitations, especially legal

XSL•FO
RenderX

framework–related limitations, ethics, and responsibilities, are less well known [5-7]. In this context of growth, understanding the relevance of synthetic data, their compliance with regulations, and their potential to address the lack of patient data in various scenarios is essential.

In this viewpoint, we provide an overview of the main applications and roles of synthetic data in cardiology, highlight the conceptual distinction between distributional and counterfactual uses, and examine the associated methodological, ethical, and regulatory challenges. We begin by providing definitions for the key terms used in the literature, and we outline the methods, benefits, and challenges of using synthetic data in cardiology research. Our central argument is that synthetic data in cardiology should not be viewed merely as a privacy-preserving technical workaround, but as a methodologically distinct form of evidence that requires dedicated governance, validation standards, and interpretability safeguards. The absence of such standards is itself a major risk: without them, synthetic data risk being adopted inconsistently, reinforcing the blind spots they are meant to address. We conclude by underlining the potential impact of synthetic data for transnational research use without compromising patient rights to privacy and by proposing 4 concrete actions for the cardiology research community to operationalize this governance agenda.

Cardiology is a particularly important test case for this argument. The high-frequency, time-series nature of key cardiac signals (electrocardiograms [ECGs], photoplethysmogram [PPG], and phonocardiograms) imposes strict fidelity requirements on generative models that are not encountered in the same way in other medical specialties. The spatial complexity of cardiac imaging modalities (magnetic resonance imaging [MRI] and computed tomography [CT]) demands higher-dimensional generative capacity, and the prevalence of rare arrhythmias and structural anomalies create chronic data scarcity for training predictive models. Additionally, the long-term disease trajectories, characteristic of chronic heart failure and coronary artery disease, challenge generative models that capture cross-sectional distributions but cannot capture longitudinal dynamics. These domain-specific constraints make the governance agenda outlined in this viewpoint particularly relevant and timely.

Clarifying the Landscape: Key Terms in Synthetic and Virtual Patient Data

Synthetic Data

The seminal work on synthetic data by the Royal Society and The Alan Turing Institute stresses functional and intentional aspects of synthetic data that have “been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s)” [8]. Synthetic data is artificial data generated from original real data and a model that is trained to reproduce the characteristics and structure of the original data [9]. Synthetic and original databases should deliver very similar results when subjected to the same statistical analysis despite different cell values. The accuracy of synthetic data as a proxy for the original data reflects the *utility* of the method and the model.

Data Augmentation

In computer vision and signal processing, augmentation typically involves transforming existing data (eg, rotation, scaling, and noise addition), while preserving labels [10]. For structured clinical or electronic health record data, augmentation more often relies on generating new patient records using statistical or machine learning (ML) models trained on real cohorts, resulting in datasets that mix real and synthetic patients [11].

Virtual Patients

Virtual patients are simulated patients used in medical training, providing a risk-free environment for health care professionals to refine their skills [6,12]. For example, a simulator was developed for patients living with heart failure to improve care delivery. It combines comprehensive hemodynamic and energetic data for personalized prognostication, such as identifying the need for heart rhythm therapy and forecasting response to device and medical therapies [13]. This is particularly useful for patients with complex clinical situations and/or drugs with a narrow therapeutic margin. However, the model can only reproduce existing elements and cannot identify new sources of interactions.

Synthetic Patients

Synthetic patients mimic real patient characteristics without using actual patient information. They are generated by statistical methods or AI and are represented in datasets, primarily for privacy-preserving research purposes [11].

Digital or Virtual Twins

Digital or virtual twins are real-time virtual replicas of physical or biological systems used to simulate patient behaviors and optimize treatments in cardiology [6,14]. These twins receive continuous updates of multiple data types to predict disease progression. Typically, twinning is partial to preserve privacy. Current innovation is increasing in cardiovascular digital twin research [15], including growing industry and academic patent activity [6,14]. Digital twins could also help plan complex surgeries, rhythm-based procedures, or to train junior cardiologists.

Taken together, these three categories exist on a spectrum: (1) virtual patients are mechanistic simulation constructs primarily used for training and procedural planning, without direct derivation from a specific individual’s records; (2) synthetic patients are population-level statistical surrogates generated to preserve privacy or augment data; and (3) digital twins are individual-level, continuously updated dynamic models. Understanding this spectrum is essential, as each category carries different validity requirements, failure modes, and regulatory implications.

How Synthetic Data Is Generated: Methods and Underlying Models

Artificial data mimicking real patient statistics can be generated from 3 sources: [11] by creating new data points similar to one

XSL•FO
RenderX

patient, by modifying data from one patient using the nearest neighbors, or by characterizing the patient population globally to create synthetic patients representing shared traits. Focusing on the last global approach, synthetic data generation can use multiple algorithms [5]. Two types of models are described in McDuff et al's review [16].

First, physical models require detailed parameterization, offering interpretability and known constraints, making them a safe choice for data generation. However, they tend to be task-specific, may not generalize well across domains, and often demand significant computational resources. Complex real-world systems can lead to model mismatches, with unknown factors posing risks to practical use.

Second, statistical generative models, such as generative adversarial networks (GANs) [17] and variational autoencoders (VAEs) [18], replicate dataset distributions and can produce realistic, high-resolution data. These models are generally easier to create than physical simulators but may lack interpretability and control. Their reliability in medical applications heavily depends on the quality of the original data (used for training) and the use of supervised learning for better regulation of output [16].

These models use explicit or implicit data generation methods [11]. Explicit approaches are used when the relationship between the variables is known or at least assumed. In these explicit approaches, it is possible to interpret the strength of association between variables (eg, covariates, outcomes, or interactions).

The first and most frequent explicit approach is mechanistic models, which are typically based on systems of ordinary or partial differential equations describing physiological processes and their interactions over time under known constraints. The second explicit approach is Bayesian networks, which can be used to generate new synthetic data through probabilistic graphical models that represent knowledge about variables and their dependencies. Bayesian networks can be used to make inferences about specific variables of interest in the network, depending on the other variables in the network.

Implicit approaches are used where there is no need to know how the variables are linked, as the model will determine this by itself, and there is no need to interpret the structure of the model. GANs and VAEs, and denoising diffusion models are all implicit approaches. GAN-based methods have significantly enhanced data analysis capabilities [11]. In health care, GANs have been successfully applied to diverse imaging modalities in cardiology, such as MRI and CT, yielding promising results [19-21]. However, a notable limitation is that most studies have relied on large training datasets or low-dimensional data. While initially less popular than GANs for data augmentation, VAEs are gaining traction in medical applications, particularly for classification and segmentation tasks [22,23]. The main limitation of VAEs is that they can produce imprecise samples when not properly tuned, especially with limited training data. Diffusion models represent an emerging class with promising fidelity for imaging tasks but remain limited in cardiology-specific applications.

To link this taxonomy to application, physical and mechanistic models are preferred when physiological interpretability and regulatory transparency are required (eg, hemodynamic simulation for device testing), at the cost of generalizability. GANs are well-suited for high-dimensional imaging data but are particularly prone to bias amplification when training datasets are small or demographically skewed. VAEs are better suited to lower-dimensional structured data and classification tasks. Bayesian networks are appropriate when variable relationships are partially known and interpretability is a priority. Choosing the right model for the intended use case is a prerequisite for the validation process described below.

Applications and Benefits for Cardiology Research

Overview

Synthetic data enables ML with large, labeled datasets that are otherwise difficult or expensive to obtain due to privacy or access constraints. Synthetic data also accelerates software testing and model pretraining and is widely used in health care to support quicker, smaller trials with fewer privacy risks. For example, synthetic data generation can reduce the time required to set up a cardiovascular data platform. In 1 reported case, synthetic data generation reduced the time required to set up a cardiovascular data platform from 1.5 years to 1 month [24], illustrating the potential for efficiency gains, though systematic evidence on typical effect sizes and the conditions under which such benefits are realized remains limited. AI models trained with synthetic or mixed data (real and synthetic data) often match or outperform those trained with real data only [25]. For example, when cancer survival prediction models were trained on synthetic data, they showed improved performance of >0.1% and, for extremely small datasets, outperformed those using only original datasets [4].

Synthetic data can significantly enhance the predictive power of AI models and optimize clinical research in cardiology through several key mechanisms.

Augmenting Data Volume and Simplifying Exportability of Health Care Datasets

Synthetic data can bolster the volume of available data, enabling more robust and comprehensive analyses [8,26]. This is particularly valuable in cardiology, where large datasets are crucial for training accurate predictive models. However, this benefit is conditional: augmenting data volume only improves downstream model performance when the generative model has successfully captured the true underlying distribution of the real data, a condition that must be verified through explicit utility and fidelity testing before the synthetic dataset is used.

Predicting Clinical Trial Results

In silico trials represent an experimental application used to expand testing methods. For example, in silico trials enable results to be extrapolated to populations with different demographics or characteristics [27], facilitating the inclusion of additional comparators, and they enable the extension of short-term study results to long-term projections [28].

XSL•FO
RenderX

Addressing Data Scarcity

When real patient data are limited, synthetic data can fill gaps and improve performance of predictive models. Synthetic data can be used to simulate rare cardiovascular conditions, allowing AI models to learn from a larger number of cases than would be available in real-world datasets [29]. For rare events, simulators can be used to create synthetic examples and thereby reduce the lack of sensitivity of a model trained with class imbalance. Examples include detecting arrhythmia from PPG or ECG [16,26].

Enhancing Data Diversity

When intentionally designed with diversity objectives and appropriate constraints, synthetic data can introduce greater variability into datasets [6]. However, this benefit is not automatic. Generative models trained on biased data will reproduce those biases. Achieving genuine diversity enhancement requires deliberate methodological choices, including stratified sampling, conditional generation targeting underrepresented groups, and rigorous subgroup validation.

Preserving Patient Privacy

By using synthetic data generated from real patient records, researchers can develop and test AI models without compromising sensitive health care–related patient information [6,24].

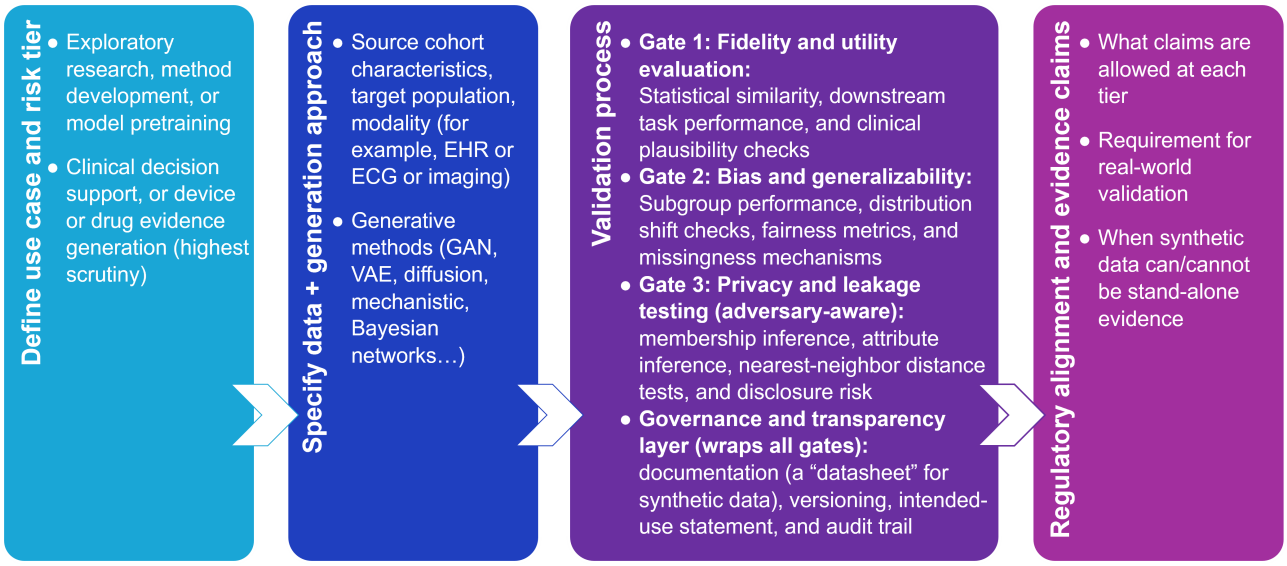
Improved Accuracy of Diagnosis

Synthetic patient data have been widely adopted for the processing of physiological signals and downstream classification in physiological measurement [16]. In particular, researchers have investigated the synthesis of cardiac signals, namely ECGs, phonocardiograms, and PPG [16]. Models trained with real ECGs augmented with synthetic ECGs have shown improved accuracy of arrhythmia diagnosis [30].

Challenges and Limitations in Applying Synthetic Data to Cardiology

While synthetic patients offer promising avenues for research and clinical applications, Figure 1 proposes a fit-for-purpose evaluation and governance workflow that links technical performance (utility and fidelity), generalizability and bias, privacy risk, and regulatory expectations to the intended use of a synthetic cohort in cardiology.

Figure 1. Proposed fit-for-purpose workflow for the evaluation and governance of synthetic cohorts. ECG: electrocardiogram; EHR: electronic health record; GAN: generative adversarial network; VAE: variational autoencoder.



Data Validity and Reliability (Fidelity)

One of the primary concerns with synthetic data is their validity. If the algorithms used to generate synthetic data do not accurately reflect the complexities of real-world patient populations, the conclusions drawn from studies using these data may be misleading. In cardiology, where patient responses can vary significantly, this is a critical issue. Addressing this systematically requires 3 complementary validation layers, as reflected in the governance workflow shown in Figure 1. First, fidelity validation involves assessing statistical similarity between synthetic and real distributions, covering both marginal and joint distributions and, for cardiology, domain-specific signal properties such as QRS morphology and RR-interval dynamics in ECG data. Second, utility validation involves testing whether models trained on synthetic data perform equivalently to those trained on real data for the downstream task of interest, using a "train-on-synthetic, test-on-real" paradigm. Third, privacy validation applies adversary-aware tests, including membership inference attacks (which test whether an adversary can identify individuals from the training set) and attribute inference evaluation (which tests whether sensitive attributes can be reconstructed), rather than relying on the absence of direct identifiers alone. All 3 layers should be treated as minimum reporting standards for any cardiology study using synthetic data.

XSL•FO
RenderX

Generalizability

Synthetic data may not generalize well to diverse or to highly specific populations [16]. Cardiology encompasses a wide range of conditions and patient demographics, and synthetic models may fail to capture the nuances of different patient groups, leading to biased results. Learned generative models produce statistics relative to the underlying data distribution, and large-parameter models have been found to amplify bias [31]. AI or ML models extract or adapt their parameters from training data, making them vulnerable to bias [32].

Ethical Considerations

The use of synthetic data raises ethical questions, particularly regarding consent and the representation of marginalized populations. Ensuring that synthetic data does not perpetuate existing biases in health care is essential for ethical research practices [7,33].

Integration With Clinical Practice

Integrating synthetic patient models into clinical workflows poses challenges. Health care providers may be hesitant to rely on models that are not validated against real-world outcomes, which can hinder the adoption of these technologies in practice.

Regulatory Blind Spots

There are currently no harmonized recommendations from agencies or regulatory authorities defining criteria for the acceptability of synthetic patient cohorts for the evaluation of health products or medical devices, and the very concept of using synthetic patient cohorts for this purpose remains without consensus [7,9,11,34,35]. By providing realistic datasets without exposing sensitive patient information, synthetic data enables health care organizations to comply with privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR); however, there is still no clear legislation specifically governing the use of synthetic data in health care. Current data protection regulations are limited in their ability to address all the potential risks associated with synthetic data, such as reidentification or data leakage [34,36]. The rise in the use of synthetic data has created an expanding industry of companies offering synthetic data solutions and enabling cross-border data sharing, sometimes beyond the confines of traditional data protection legislation. Regulatory agencies are actively exploring the use of synthetic data, but as of 2025, synthetic data is not yet broadly trusted as stand-alone evidence for regulatory approval, and findings derived from synthetic data are required to be validated on real data.

Existing partial frameworks offer some orientation but leave critical gaps. The European Medicines Agency's 2024 *Reflection Paper on the Use of Artificial Intelligence (AI) in the medicinal product lifecycle* [32] acknowledges simulation and synthetic data as emerging tools but stops short of defining acceptability criteria for their use as stand-alone evidence. The Food and Drug Administration's evolving guidance on AI and ML-based software as a medical device similarly does not yet address synthetic patient cohorts as a distinct evidence category [37]. Any future harmonized framework would need to specify at minimum: (1) the intended-use scope within which a synthetic dataset has been validated and may legitimately be used, (2) minimum fidelity thresholds appropriate to the application risk tier, (3) mandatory adversary-aware privacy testing before regulatory submission, and (4) a requirement that findings derived from synthetic data be validated on real-world data before supporting regulatory decisions. Until such a framework exists, researchers and sponsors should proactively document their datasets against these criteria to support transparency and anticipate future expectations. Recent work underscores that the core barrier is not only regulatory uncertainty but also the lack of agreed, quantitative ways to demonstrate that a synthetic dataset is both useful and safe. In their scoping review, Kaabachi et al [5] identified dozens of distinct approaches to assess synthetic data utility and only a small set of approaches used to evaluate privacy risk, concluding that many studies do not assess privacy protection at all and that, when they do, residual risks are often underestimated. This supports the call for harmonized evaluation frameworks that pair utility metrics with explicit, adversary-aware privacy testing before synthetic cohorts are considered acceptable evidence in cardiology research or product evaluation.

Privacy

Although synthetic data theoretically offer patient privacy protection, there remains a risk that synthetic data generation may leak personal information [8,16]. Strictly, it is not possible to determine whether synthetic data are sufficiently different from the original samples to be considered anonymized [36]. There are no robust and objective methods of determining whether a synthetic dataset is sufficiently different from the original real dataset to be classified as truly anonymous, and recent reviews highlight inconsistent practice and the frequent absence of privacy-risk evaluation in published synthetic data studies [5]. This is a fundamental, not peripheral, problem: without a reliable method to certify sufficient dissimilarity from training data, the use of synthetic datasets in research settings remains legally ambiguous under both GDPR and HIPAA, regardless of scientific utility. In practice, privacy assurance requires adversary-aware testing rather than reliance on the absence of direct identifiers. Two tests are particularly relevant: membership inference attacks, which evaluate whether an adversary can determine that a specific individual's record was used in training; and attribute inference evaluation, which tests whether sensitive attributes (such as diagnosis or treatment history) can be reconstructed from synthetic records. Kaabachi et al [5] found that most published synthetic data studies do not perform privacy-risk evaluation at all, and that when they do, adversarial attack methods are rarely applied and residual risks are frequently underestimated. The cardiology field should treat adversary-aware privacy validation as a nonnegotiable component of any synthetic data reporting standard, consistent with the privacy gate in Figure 1.

Technical Limitations

Algorithms used to generate synthetic data can have limitations in terms of complexity and accuracy. In cardiology, where patient data can be highly variable and influenced by numerous factors, these limitations can impact the effectiveness of synthetic data for modeling patient outcomes.

XSL•FO
RenderX

Conclusions

Overall, while synthetic patient research and synthetic data hold substantial promise for advancing cardiovascular research, careful consideration of the associated challenges is necessary to ensure that these tools are used effectively in clinical practice. Synthetic datasets and in silico studies could transform the future of cardiovascular research, allowing the use and transfer of reliable datasets without compromising patient privacy and addressing data scarcity. Realizing this potential depends on recognizing and governing the 2 fundamentally different roles that synthetic data can play, either as a privacy-preserving distributional substitute for existing data or as a counterfactual simulator for rare events and hypothetical scenarios, and on applying validation criteria appropriate to each role. Within strict boundaries and under specific conditions to limit potential bias, synthetic data can uphold privacy, equity, and safety, but care must be taken to avoid introducing flaws, blind spots, and propagating or exaggerating biases. Synthetic datasets represent a fundamentally different form of evidence, and their value depends on how rigorously their limitations, sources of bias, and intended uses are defined and governed. In the absence of shared methodological standards and explicit regulatory expectations, synthetic data risks being adopted inconsistently, potentially reinforcing blind spots rather than mitigating them. We argue that progress in this field now requires a collective shift from exploring what synthetic data *can* do to agreeing on when, how, and for which purposes they *should* be used in cardiology research. Within clearly defined boundaries and under transparent validation and governance conditions, synthetic data can support innovation, data sharing, and equity without compromising scientific credibility or patient trust. Importantly, the challenges associated with synthetic data are amplified in cardiology, where clinically relevant information is often encoded in high-frequency signals, complex cardiac anatomy, and long-term disease trajectories.

To operationalize this governance agenda, we propose 4 concrete actions for the cardiology research community. First, mandatory three-layer validation for any publication using synthetic cardiology data: (1) fidelity validation (statistical similarity, including signal-specific properties such as QRS morphology for ECG data), (2) utility validation (train-on-synthetic, test-on-real performance equivalence for the downstream task), and (3) privacy validation through adversary-aware testing, including membership inference and attribute inference evaluation. Second, mandatory subgroup reporting requires performance metrics should be disaggregated by relevant demographic subgroups, rather than reported only at the population level. Third, intended-use scoping dictating that any synthetic dataset should document the specific purpose for which it was validated, consistent with the fit-for-purpose principle illustrated in Figure 1. Fourth, engagement by cardiology societies and regulatory agencies to define domain-specific acceptability thresholds for synthetic evidence, beginning with lower-stakes applications such as model pretraining and simulation before moving to higher-stakes regulatory submissions.

Acknowledgments

The authors thank Pippa McKelvie-Sebileau, PhD, from Horiana for editing and proofreading. AI tools (Claude) were used to suggest language improvements within the manuscript.

Funding

No external funding was received for this study.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Authors' Contributions

Conceptualization: LB, SM, SL, PJ
Investigation: LB, SL, JJ, PJ
Methodology: LB, SM, VB
Supervision: PJ
Validation: SM
Visualization: LB
Writing—original draft: LB
Writing—review and editing: LB, SM, SL, JJ, VB, PJ
All authors approved the final version for submission.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Rujas M, Martín Gómez Del Moral Herranz R, Fico G, Merino-Barbancho B. Synthetic data generation in healthcare: a scoping review of reviews on domains, motivations, and future applications. Int J Med Inform 2025 Mar;195:105763. [doi: 10.1016/j.ijmedinf.2024.105763] [Medline: 39719743]
2. Kokosi T, Harron K. Synthetic data in medical research. BMJ Med 2022;1(1):e000167. [doi: 10.1136/bmjmed-2022-000167] [Medline: 36936569]
3. Gonzales A, Guruswamy G, Smith SR. Synthetic data in health care: a narrative review. PLOS Digit Health 2023 Jan;2(1):e0000082. [doi: 10.1371/journal.pdig.0000082] [Medline: 36812604]
4. Kim H, Jang WS, Sim WS, et al. Synthetic data improve survival status prediction models in early-onset colorectal cancer. JCO Clin Cancer Inform 2024 Jan;8:e2300201. [doi: 10.1200/CCI.23.00201] [Medline: 38271642]
5. Kaabachi B, Despraz J, Meurers T, et al. A scoping review of privacy and utility metrics in medical synthetic data. NPJ Digit Med 2025 Jan 27;8(1):60. [doi: 10.1038/s41746-024-01359-3] [Medline: 39870798]
6. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. NPJ Digit Med 2023 Oct 9;6(1):186. [doi: 10.1038/s41746-023-00927-3] [Medline: 37813960]
7. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 6;6(1):120. [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]
8. Jordon J, Szpruch L, Houssiau F, et al. Synthetic data - what, why and how? : The Royal Society; 2022 URL: https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic_Data_Survey-24.pdf [accessed 2026-07-18]
9. Synthetic data. European Data Protection Supervisor. URL: https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en [accessed 2026-07-18]
10. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6:60. [doi: 10.1186/s40537-019-0197-0]
11. Allassonniere S, Fraysse JL. Données de santé artificielles: analyse et pistes de réflexion [Article in French]. : BOTdesign; 2024 URL: https://static.botdesign.net/docs/VF_Livre_blanc_Donn%C3%A9es_de_sant%C3%A9_artificielles-250424.pdf [accessed 2026-07-18]
12. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
13. Grinstein J, Blanco PJ, Torii R, Goswami R. The virtual patient simulator: using hemodynamics and the cardioenergetic profile to optimize heart failure care. JACC Basic Transl Sci 2025 Mar;10(3):259-261. [doi: 10.1016/j.jacbts.2025.01.009] [Medline: 40139863]
14. Venkatesh KP, Raza MM, Kvedar JC. Health digital twins as tools for precision medicine: considerations for computation, implementation, and regulation. NPJ Digit Med 2022 Sep 22;5(1):150. [doi: 10.1038/s41746-022-00694-7] [Medline: 36138125]
15. Coorey G, Figtree GA, Fletcher DF, et al. The health digital twin to tackle cardiovascular disease-a review of an emerging interdisciplinary field. NPJ Digit Med 2022 Aug 26;5(1):126. [doi: 10.1038/s41746-022-00640-7] [Medline: 36028526]
16. McDuff D, Curran T, Kadambi A. Synthetic data in healthcare. arXiv. Preprint posted online on Apr 6, 2023. [doi: 10.48550/arXiv.2304.03243]
17. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. arXiv. Preprint posted online on Jun 10, 2014. [doi: 10.48550/arXiv.1406.2661]
18. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv. Preprint posted online on Dec 20, 2013. [doi: 10.48550/arXiv.1312.6114]
19. Peirlinck M, Costabal FS, Yao J, et al. Precision medicine in human heart modeling: perspectives, challenges, and opportunities. Biomech Model Mechanobiol 2021 Jun;20(3):803-831. [doi: 10.1007/s10237-021-01421-z] [Medline: 33580313]
20. Merle M, Collot F, Castelneau J, et al. MUSIC: cardiac imaging, modelling and visualisation software for diagnosis and therapy. Appl Sci 2022;12(12):6145. [doi: 10.3390/app12126145]
21. Nakatani Y, Nuñez-Garcia M, Cheniti G, et al. Preoperative personalization of atrial fibrillation ablation strategy to prevent esophageal injury: impact of changes in esophageal position. J Cardiovasc Electrophysiol 2022 May;33(5):908-916. [doi: 10.1111/jce.15447] [Medline: 35274776]
22. Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc 2019 Mar 1;26(3):228-241. [doi: 10.1093/jamia/ocy142] [Medline: 30535151]
23. Diaz-Pinto A, Alle S, Nath V, et al. MONAI Label: a framework for AI-assisted interactive labeling of 3D medical images. Med Image Anal 2024 Jul;95:103207. [doi: 10.1016/j.media.2024.103207] [Medline: 38776843]
24. Azizi Z, Lindner S, Shiba Y, et al. A comparison of synthetic data generation and federated analysis for enabling international evaluations of cardiovascular health. Sci Rep 2023 Jul 17;13(1):11540. [doi: 10.1038/s41598-023-38457-3] [Medline: 37460705]
25. McDuff D, Hernández J, Wood E, Liu X, Baltrusaitis T. Advancing non-contact vital sign measurement using synthetic avatars. arXiv. Preprint posted online on Oct 24, 2020. [doi: 10.48550/arXiv.2010.12949]

XSL•FO
RenderX

26. Williams MC, MacArthur JA, Forsyth R, Petersen SE. New ways to use imaging data in cardiovascular research: survey of opinions on federated learning and synthetic data. Eur Heart J Imaging Methods Pract 2025 Jan;3(1):qyaf012. [doi: 10.1093/ehjimp/qyaf012] [Medline: 40066149]
27. Myles P, Ordish J, Tucker A. The potential synergies between synthetic data and in silico trials in relation to generating representative virtual population cohorts. Prog Biomed Eng 2023;5:013001. [doi: 10.1088/2516-1091/acafbf]
28. Angoulvant D, Granjeon-Noriot S, Amarenco P, et al. In-silico trial emulation to predict the cardiovascular protection of new lipid-lowering drugs: an illustration through the design of the SIRIUS programme. Eur J Prev Cardiol 2024 Nov 11;31(15):1820-1830. [doi: 10.1093/eurjpc/zwae254] [Medline: 39101472]
29. Lutfi H, Spittler T, Ibrahim HM. Using synthetic data and machine learning to predict heart rate and enable lifestyle recommendations. J Med Artif Intell 2024 Sep;7. [doi: 10.21037/jmai-24-35]
30. Skandarani Y, Lalande A, Afilalo J, Jodoin PM. Generative adversarial networks in cardiology. Can J Cardiol 2022 Feb;38(2):196-203. [doi: 10.1016/j.cjca.2021.11.003] [Medline: 34780990]
31. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? In: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency: Association for Computing Machinery; 2021:610-623. [doi: 10.1145/3442188.3445922]
32. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. European Medicines Agency. 2024. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf [accessed 2025-02-28]
33. Boraschi D, van der Schaar M, Costa A, Milne R. Governing synthetic data in medical research: the time is now. Lancet Digit Health 2025 Apr;7(4):e233-e234. [doi: 10.1016/j.landig.2025.01.012] [Medline: 39984419]
34. Hendricks-Sturrup R, Nafie M. 64 Synthetic data generation using generative AI to support biomedical innovation: a health policy perspective. J Clin Transl Sci 2026;10(s1):20. [doi: 10.1017/cts.2026.10290]
35. Foraker R, Morrow JD, Johnson JA, Wilcox AB, Forster AJ, Payne PR. Understanding synthetic data: artificial datasets for real-world evidence. BMJ Evid Based Med 2026 May 21;31(3):148-151. [doi: 10.1136/bmjebm-2024-113617] [Medline: 40602833]
36. Arora A, Arora A. Synthetic patient data in health care: a widening legal loophole. Lancet 2022 Apr 23;399(10335):1601-1602. [doi: 10.1016/S0140-6736(22)00232-X] [Medline: 35358423]
37. Artificial intelligence-enabled device software functions: lifecycle management and marketing submission recommendations. U.S. Food & Drug Administration. 2025. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing [accessed 2026-07-08]

Abbreviations

CT: computed tomography
ECG: electrocardiogram
GAN: generative adversarial network
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
ML: machine learning
MRI: magnetic resonance imaging
PPG: photoplethysmogram
VAE: variational autoencoder

Edited by A Coristine; submitted 05.Feb.2026; peer-reviewed by AM Nandialath, S Guan; revised version received 18.Jun.2026; accepted 19.Jun.2026; published 31.Jul.2026.

Please cite as:
Baschet L, Marque S, Locret S, Jenssen J, Barbet V, Jourdain P
Addressing the Challenges in Using Synthetic Data for Health Research: Application to Cardiology
JMIR Cardio 2026;10:e92930
URL: https://cardio.jmir.org/2026/1/e92930
doi:10.2196/92930

© Louise Baschet, Sebastien Marque, Stéphane Locret, Jens Jenssen, Vanessa Barbet, Patrick Jourdain. Originally published in JMIR Cardio (https://cardio.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Large Language Models in Cardiology: Systematic Review

Moran Gendler[1*], MD; Girish N Nadkarni[2], Dr med; Karin Sudri[3], MA; Michal Cohen-Shelly[3], MBA; Benjamin S Glicksberg[2], PhD; Orly Efros[4,5], MD; Shelly Soffer[4,6*], MD; Eyal Klang[2*], MD

[1]Azrieli Faculty of Medicine, Bar-Ilan University, Henrietta Szold St 8, Safed, Israel, Safed, Israel

[2]Windreich Department of AI and Human Health, Mount Sinai Medical Center, Mount Sinai, New York, NY, United States

[3]Sagol AI Hub, ARC Innovation Center, Sheba Medical Center, Ramat Gan, Israel

[4]School of Medicine, Tel Aviv University, Tel Aviv, Israel

[5]National Hemophilia Center and Thrombosis Institute, Sheba Medical Center, Ramat Gan, Israel

[6]Institute of Hematology, Davidoff Cancer Center, Rabin Medical Center, Petah-Tikva, Israel

[*]these authors contributed equally

Corresponding Author:
Moran Gendler, MD
Azrieli Faculty of Medicine, Bar-Ilan University, Henrietta Szold St 8, Safed, Israel, Safed, Israel

Abstract

Background: Large language models (LLMs) are increasingly used in health care, but their role in cardiology has not yet been systematically evaluated.

Objective: This review aimed to assess the applications, performance, and limitations of LLMs across diverse cardiology tasks, including chronic and progressive conditions, acute events, education, and diagnostic testing.

Methods: A systematic search was conducted in PubMed and Scopus for studies published up to April 14, 2024, using keywords related to LLMs and cardiology. Studies evaluating LLM outputs in cardiology-related tasks were included. Data were extracted across 5 predefined domains and the risk of bias was assessed using an adapted QUADAS-2 tool (developed by Whiting et al at the University of Bristol). The review protocol was registered in PROSPERO (CRD42024556397).

Results: A total of 33 studies contributed quantitative outcome data to a descriptive synthesis. Across chronic conditions, ChatGPT-3.5 (OpenAI) answered 91% (43/47) heart failure questions accurately, although readability often required college-level comprehension. In acute scenarios, Bing Chat omitted key myocardial infarction first aid steps in 25% (5/20) to 45% (9/20) of cases, while cardiac arrest information was rated highly (mean 4.3/5, SD 0.7) but written above recommended reading levels. In physician education tasks, ChatGPT-4 (OpenAI) demonstrated higher accuracy than ChatGPT-3.5, improving from 38% (33/88) to 66% (58/88). In patient education studies, ChatGPT-4 provided scientifically adequate explanations (5.0 - 6.0/7) comparable to hospital materials but at higher reading levels (11th vs 7th grade). In diagnostic testing, ChatGPT-4 interpreted 91% (36/40) electrocardiogram vignettes correctly, significantly better than emergency physicians (31/40, 77%; $P< .001$), but showed lower performance in echocardiography.

Conclusions: LLMs show meaningful potential in cardiology, especially for education and electrocardiogram interpretation, but performance varies across clinical tasks. Limitations in emergency guidance and readability, as well as small in silico study designs, highlight the need for multimodal models and prospective validation.

Trial Registration: PROSPERO CRD42024556397; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024556397

(JMIR Cardio 2026;10:e76734) doi:10.2196/76734

KEYWORDS

artificial intelligence; natural language processing; large language models; generative AI; LLMs; cardiology

Introduction

Large language models (LLMs) such as OpenAI's ChatGPT, Google's Gemini, and Meta's LLaMA are advancing natural language processing by generating, understanding, and interpreting text. These models process text to produce coherent responses, understand context, summarize information, and engage in conversations [1]. Their application in health care, particularly in cardiology, offers significant benefits due to their ability to analyze diverse and complex data—from patient records to imaging studies [2,3].

In cardiology, LLMs are increasingly being used to assist in the management of cardiovascular diseases by organizing and making clinical data more accessible [4]. These models can enhance diagnostic accuracy, personalize treatment plans, and identify patterns in large datasets that traditional methods might

XSL•FO
RenderX

overlook [5,6]. Additionally, LLMs offer the potential to automate routine documentation, thereby reducing the administrative burden on health care providers [7,8]. However, integrating LLMs into clinical workflows poses challenges, and effective implementation is crucial to realizing their potential to improve patient care in cardiology [8-10].

Recent reviews have highlighted the emerging role of LLMs in cardiology. Sharma et al [9] provided an early synthesis of ChatGPT applications, focusing on health literacy, clinical care, and research up to September 2023. Boonstra et al [7] more recently examined LLMs across cardiovascular disease, with emphasis on prevention and patient education. Our review complements these works by incorporating a broader search, applying PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 methodology and organizing findings into 5 clinically relevant domains. It emphasizes current uses of LLMs in cardiology, their potential impact on care and patient outcomes, and the barriers to their practical application.

Methods

Overview

This review was conducted according to the PRISMA guidelines [10] (Checklist 1).

Search Strategy

A comprehensive literature search was conducted to identify studies on the application of LLMs in cardiology. The search was performed on April 14, 2024, in PubMed and Scopus, using a combination of keywords and Medical Subject Headings (MeSH) related to both cardiology and LLMs. The cardiology terms included “Echocardiography,” “Arrhythmias,” “Cardiac Output,” “Heart Failure,” “Heart Valve Diseases,” “Myocardial Ischemia,” “Acute Coronary Syndrome,” and “Electrocardiogram.” The LLM terms included “ChatGPT,” “Large Language Models,” “OpenAI,” “Microsoft Bing Chat,” “Google Bard” and “Google Gemini.” In Scopus, searches were conducted using the TITLE ABS KEY field to ensure consistency across databases. Scopus was included alongside PubMed to broaden coverage, capturing interdisciplinary studies at the intersection of artificial intelligence and cardiology that may not be indexed in PubMed. The complete search strategies are available in Multimedia Appendix 1. This review was registered with PROSPERO (CRD42024556397) [11].

Study Selection

We included studies that (1) evaluated an application of LLMs in a specific field within cardiology, (2) were published in English, and (3) were peer reviewed. In addition to full original research articles, short reports and letters containing original data or quantitative analyses were also eligible. Studies that were non–LLM-related, non–cardiology-focused, or purely conceptual without empirical evaluation were excluded. Abstracts, conference papers, critical letters, and editorial commentaries were also excluded.

The search was supplemented by manual screening of the reference lists of included studies. Two reviewers (MG and SS) independently screened the titles and abstracts to determine whether the studies met the inclusion criteria. Full-text articles were reviewed when the title met the inclusion criteria or when there was any uncertainty. Disagreements were resolved by a third reviewer (EK).

Data Extraction

Two independent reviewers (MG and SS) extracted data from the included studies using a standardized data extraction form. Discrepancies were resolved through discussion or consultation with a third reviewer (EK). Extracted information included study design, sample size, LLM application details (eg, LLM features examined, assessment method, validation metrics, and reference guidelines used for accuracy comparison), main findings, and limitations.

Quality Assessment

Risk of bias was assessed using the QUADAS-2 tool (developed by Whiting et al at the University of Bristol) [12], which is widely applied in diagnostic accuracy research. This framework was selected because many included studies evaluated LLMs in diagnostic or decision-making roles, making QUADAS-2 particularly suitable for systematically assessing potential bias in study design, case selection, index test conduct, and reference standards. A detailed summary of the assessments is presented in Table S1 in Multimedia Appendix 1.

Since the studies evaluated LLM performance rather than human diagnostics, several adaptations were applied. In the patient selection domain, we assessed the transparency and representativeness of the test cases used to evaluate LLMs. Bias in this domain was considered high when studies used unreported or simplified cases that did not reflect real-world clinical variability. In the index test domain, we evaluated the standardization of prompts and scoring (number of runs and grading rules), model transparency (version, release date, and parameters), blinding to the reference standard during testing, and avoidance of post hoc prompt modification or selective reporting. The reference standard domain was adapted to assess the reliability of the comparator or ground truth (eg, expert consensus, guideline-based answers, or validated datasets).

Data Synthesis

A narrative synthesis of the findings from the included studies was conducted. Due to anticipated heterogeneity in study designs and outcomes, a meta-analysis was not planned. Instead, the focus was on summarizing the applications, benefits, and limitations of LLMs in cardiology as reported in the included studies and identifying areas for future research. In this paper, “quantitative synthesis” refers to descriptive reporting of numerical outcomes extracted from individual studies (eg, accuracy rates, agreement statistics, readability scores, and error frequencies) without statistical pooling or calculation of combined effect estimates.

To structure the analysis, included studies were grouped into 5 categories that reflected the main areas of LLM application in cardiology. Two reviewers (MG and SS) independently categorized the studies, and any discrepancies were resolved through discussion. In cases where consensus could not be reached, a final decision was made by EK. Data were extracted

XSL•FO
RenderX

for each category on study objectives, type of task, LLMs assessed, evaluation methods, and key performance outcomes.

Chronic and Progressive Cardiac Conditions

Studies were included if they assessed LLMs in long-term cardiac conditions such as heart failure, hypertension, valvular disease, or atrial fibrillation.

Acute Cardiac Events

This group included studies evaluating LLMs in acute scenarios, including resuscitation, cardiac arrest, and myocardial infarction.

Physician Education

Studies were categorized here if they tested LLMs on cardiology training, examination-style questions, or case vignettes aimed at medical professionals. Studies that focused on physician clinical decision-making and compared it with LLM performance were also included under this group.

Patient Education

This group covered studies where LLMs provided information or educational content directly to patients.

Cardiac Diagnostics Tests

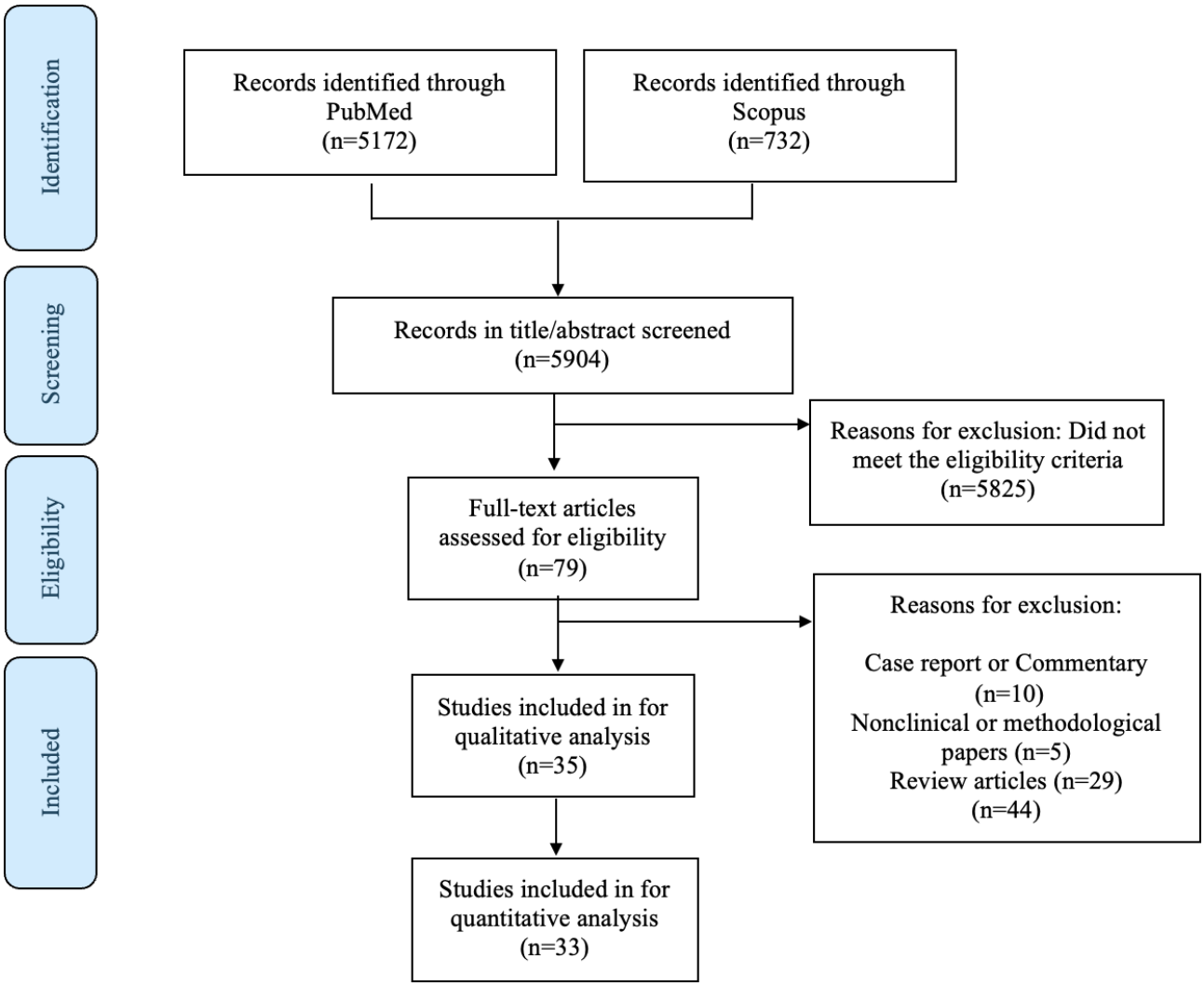
Studies were included if they examined the use of LLMs for diagnostic interpretation, such as electrocardiograms (ECGs), echocardiography, and cardiac imaging.

Results

Overview

A total of 35 articles were identified for inclusion [5,13-46]. Of these, 3 were retrieved exclusively from PubMed [27,31,39], while the remaining articles were identified in both databases. Following full-text assessment, 33 [5,13-30,33-46] studies contributed quantitative outcome data to the descriptive quantitative synthesis [5,13-30,33-46], whereas 2 studies were included in qualitative synthesis only and were therefore not included in the quantitative synthesis [31,32]. These 2 studies reported conceptual analyses and descriptive observations rather than measurable performance metrics (Figure 1).

Figure 1. Flow diagram of the search and inclusion process.



General details about the included articles, descriptions of their characteristics, main outcomes, and their advantages and limitations are summarized in Tables 1-4, respectively. Tables S1 and S2 in Multimedia Appendix 1 provide additional detail on characteristics and outcomes. Table S3 in Multimedia Appendix 1 provides a detailed evaluation of each article using

XSL•FO
RenderX

the QUADAS-2 tool. Figure 2 shows the categorization of articles into core groups with corresponding cardiology subfields.

XSL•FO
RenderX

Table . Studies included.

Category and study	Publication date	Title	Journal	PMID
Chronic and progressive cardiac conditions				
Dimitriadis et al [27]	2024 March	ChatGPT and patients with heart failure	Angiology	38451243
Riddle et al [33]	2023 December	College-level reading is required to understand ChatGPT's answers to lay questions relating to heart failure	European Journal of Heart Failure	37964183
Krittanawong et al [34]	2023 November and December	Assessing the potential of ChatGPT for patient education in the cardiology clinic	Progress in Cardiovascular Diseases	37832625
Rouhi et al [35]	2024 March	Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study	Cardiology and Therapy	38194058
Hillmann et al [36]	2023 December	Accuracy and comprehensibility of chat-based artificial intelligence for patient information on atrial fibrillation and cardiac implantable electronic devices	EP Europace	38127304
Van Bulck et al [37]	2024 January	What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions	European Journal of Cardiovascular Nursing	37094282
Kassab et al [25]	2023 November	Comparative analysis of chat-based artificial intelligence models in addressing common and challenging valvular heart disease clinical scenarios	Journal of the American Heart Association	37982246
Han et al [24]	2024 January	Evaluation of GPT-4 for 10-year cardiovascular risk prediction- insights from the UK Biobank and KoGES[a] data	iScience	38357664
Ali et al [23]	2023 November	Mapping the heartbeat of America with ChatGPT-4- unpacking the interplay of social vulnerability, digital literacy, and cardiovascular mortality in county residency choices	Journal of Personalized Medicine	38138852
Li et al [22]	2024 March	Potential multidisciplinary use of large language models for addressing queries in cardio-oncology	Journal of the American Heart Association	38497458
Yano et al [21]	2023 November	Relevance of ChatGPT's responses to common hypertension-related patient inquiries	Hypertension	37916418

XSL•FO
RenderX

Category and study	Publication date	Title	Journal	PMID
Kusunose et al [46]	2023 June	Evaluation of the accuracy of ChatGPT in answering clinical questions on the Japanese Society of Hypertension Guidelines	Circulation Journal	37286486
Al Tibi et al [45]	2024 March	A retrospective comparison of medication recommendations between a cardiologist and ChatGPT-4 for hypertension patients in a rural clinic	Cureus	38586651
Acute cardiac events				
Birkun and Gautam [38]	2024 January	Large language model-based chatbot as a source of advice on first aid in heart attack	Current Problems in Cardiology	37640177
Scquizzato et al [39]	2024 January	Testing ChatGPT ability to answer laypeople questions about cardiac arrest and cardiopulmonary resuscitation	Resuscitation	38081504
Safranek et al [40]	2024 March	Automated HEART[b] score determination via ChatGPT: Honing a framework for iterative prompt development	Journal of the American College of Emergency Physicians Open	38481520
Birkun [43]	2023 August	Performance of an artificial intelligence-based chatbot when acting as EMS[c] dispatcher in a cardiac arrest scenario	Internal and Emergency Medicine	37603142
Physician education				
Harskamp and De Clercq [41]	2024 February	Performance of ChatGPT as an AI[d]-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2)	Acta Cardiologica	38348835
Skalidis and Cagnina [16]	2024 April	ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story?	European Heart Journal - Digital health	37265864
Yavuz and Kahraman [15]	2024 March	Evaluation of the prediagnosis and management of ChatGPT-4.0 in clinical cases in cardiology	Future Cardiol	39049771
Gritti et al [13]	2024 February	Progression of an artificial intelligence Chatbot (ChatGPT) for pediatric cardiology educational knowledge assessment	Pediatric Cardiology	38170274
Lee et al [5]	2023 October	Evaluating the clinical decision-making ability of large language models using MKSAP-19[e] cardiology questions	JACC Advance	38938709
Patient education				

XSL•FO
RenderX

Category and study	Publication date	Title	Journal	PMID
Günay et al [19]	2024 March	AI in patient education: assessing the impact of ChatGPT-4 on conveying comprehensive information about chest pain	The American Journal of Emergency Medicine	38242775
Bushuven et al [20]	2023 November	"ChatGPT, Can You Help Me Save My Child's Life?" - diagnostic accuracy and supportive capabilities to lay rescuers by ChatGPT in prehospital basic life support and pediatric advanced life support cases - an in-silico analysis	Journal of Medical Systems	37987870
Lautrup et al [18]	2023 November	Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice	Open Heart	37945282
Moons and Van Bulck [17]	2024 March	Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept.	European Journal of Cardiovascular Nursing	37603843
Almagazzachi et al [44]	2024 February	Generative artificial intelligence in patient education: ChatGPT takes on hypertension questions	Cureus	38435177
Cardiac diagnostic tests				
Fijačko et al [30]	2023 December	Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images?	Resuscitation	37884222
Zhu et al [42]	2023 July	ChatGPT can pass the AHA[f] exams: open-ended questions outperform multiple-choice format	Resuscitation	37349064
King et al [14]	2024 February	GPT-4V passes the BLS[g] and ACLS[h] examinations: an analysis of GPT-4V's image recognition capabilities	Resuscitation	38160904
Günay et al [29]	2024 March	Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment	The American Journal of Emergency Medicine	38507847
Kangiszer et al [28]	2024 March	Low performance of ChatGPT on echocardiography board review questions	JACC: Cardiovascular Imaging	37943230

XSL•FO
RenderX

Category and study	Publication date	Title	Journal	PMID
Sarangi et al [26]	2023 December	Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: perspectives of four large language models	The Indian Journal of Radiology & Imaging	38549881

[a]KoGES: Korean Genome and Epidemiology Study.

[b]HEART: History, ECG, Age, Risk factors, Troponin risk algorithm.

[c]EMS: Emergency Medical Services.

[d]AI: artificial intelligence.

[e]MKSAP-19: Medical Knowledge Self-Assessment Program (19th edition).

[f]AHA: American Heart Association.

[g]BLS: Basic Life Support.

[h]ACLS: Advanced Cardiovascular Life Support.

XSL•FO
RenderX

Table . Study characteristics.

Category and authors	Tasks	LLM[a] features examined	LLM examined
Chronic and progressive cardiac conditions			
Dimitriadis et al [27]	Answering questions about the management of HF[b].	• Accuracy	• ChatGPT- 3.5
Riddell et al [33]	Answer a set of hypothetical queries from a patient with HFrEF[c].	• Readability	• ChatGPT- 4
Krittanawong et al [34]	Answer HF-related questions.	• Reliability	• ChatGPT- 3.5
Rouhi et al [35]	Rewrite patient education materials to meet recommended reading skill levels for patients with AS[d].	• Readability • Simplification	• Google Bard • ChatGPT- 3.5
Hillmann et al [36]	Answering questions about AF[e] and CIED[f].	• Readability • Appropriateness • Comprehensibility	• Google Bard • Bing Chat • ChatGPT-4
Van Bulck and Moons [37]	Responses to virtual prompts by patients (CHD[g], AF, HF, and Chol[h]).	• Trustworthy • Valuable • Dangerous	• ChatGPT-3
Kassab et al [25]	Answer correctly on 15 patient-centered and 15 physician-centered VHD[i] queries.	• Accuracy	• ChatGPT-4 • Google Bard
Han et al [24]	Predict 10-year CVD[j] risk from cohort data.	• Accuracy • Robustness	• ChatGPT- 3.5 • ChatGPT-4
Ali et al [23]	Predicting age-adjusted cardiovascular mortality across 3118 US counties and identifying associations with social vulnerability and digital literacy indices.	• Assist with regression modeling • Code generation	• ChatGPT-4
Li et al [22]	Answer 25 guideline-based cardio-oncology queries.	• Accuracy • Guideline adherence	• ChatGPT- 3.5 • ChatGPT-4 • Google Bard • Llama2 • Claude2
Yano et al [21]	Answer 20 common hypertension FAQs[k] in Japanese and English.	• Appropriateness • Language consistency	• ChatGPT-4
Kusunose et al [46]	Answering guideline-based clinical questions on hypertension.	• Accuracy	• ChatGPT- 3.5
Al Tibi et al [45]	Comparing antihypertensive medication recommendations between ChatGPT-4 and a cardiologist using real-world patient data.	• Accuracy	• ChatGPT-4
Acute cardiac events			
Birkun and Gautam [38]	Examination of the ability of chatbots to guide first aid for heart attacks.	• Guideline adherence • Readability	• Bing Chat
Scquizzato et al [39]	Answer lay FAQs on cardiac arrest and CPR[l].	• Accuracy • Readability	• ChatGPT-3.5
Safranek et al [40]	Extract data from notes and compute HEART[m] score; test iterative prompt framework.	• Accuracy • Guideline adherence	• ChatGPT- 3.5 • ChatGPT-4

XSL•FO
RenderX

Category and authors	Tasks	LLM[a] features examined	LLM examined
Birkun [43]	Evaluate the ability to operate as an automated assistant for recognition of cardiac arrest and real-time CPR instructions.	• Accuracy • Guideline adherence	• New Bing chatbot
Physician education			
Harskamp and De Clercq [41]	Answering questions related to common cardiac symptoms or conditions.	• Accuracy	• ChatGPT- 3.5 January 2023 • ChatGPT-3.5 September 2023 version
Skalidis et al [16]	Test ChatGPT on exam-style questions.	• Accuracy	• ChatGPT-3
Yavuz and Kahraman [15]	Assess ChatGPT-4 inthediagnosis and management of cases.	• Accuracy • Difficulty	• ChatGPT-4
Gritti et al [13]	Compare ChatGPT-4 and ChatGPT-3.5 in accuracy on multiple-choice pediatric cardiology questions.	• Accuracy	• ChatGPT 3.5 • ChatGPT-4
Lee et al [5]	Achieving a passing score of 50 % using MKSAP-19's[n] cardiology questions.	• Accuracy	• ChatGPT- 3.5 • ChatGPT-4 • PubMedGPT
Patient education			
Günay et al [19]	Compare ChatGPT-4 vs hospital websites on chest pain FAQs.	• Readability • Guideline adherence	• ChatGPT-4
Bushuven et al [20]	Evaluate ChatGPT diagnostic and capabilities in BLS[o] and PALS[p] scenarios.	• Accuracy	• ChatGPT-3.5 • ChatGPT-4
Lautrup et al [18]	Respond to prompts on 4 cardiovascular topics (MI[q], PAD[r], VV[s], and CP)[t].	• Accuracy	• ChatGPT-4
Moons and Van Bulck [17]	Simplify patient info from journals.	• Readability	• ChatGPT • Google Bards
Almagazzachi et al [44]	Answering hypertension questions and assessing reproducibility across repeated runs.	• Accuracy	• ChatGPT (version not specified)
Cardiac diagnostic tests			

XSL•FO
RenderX

Category and authors	Tasks	LLM[a] features examined	LLM examined
Fijačko et al [30]	Interpreting ECG images.	• Accuracy	• Google Bard • Bing Chat • ChatGPT-4
Zhu et al [42]	Interpreting ECG images.	• Accuracy	• ChatGPT
King et al [14]	Interpreting ECG images.	• Accuracy	• ChatGPT-4V
Günay et al [29]	Interpreting ECG data.	• Accuracy	• ChatGPT-4
Kangiszer et al [28]	Answer correctly echocardiography board review questions and provide explanations that reflect standards of practice.	• Accuracy	• ChatGPT-4
Sarangi et al [26]	Generate DDx for 15cardiac or thoracic imaging patterns.	• Diagnostic accuracy	• ChatGPT- 3.5 • Google Bard • Microsoft Bing Perplexity

[a]LLM: large language model.
[b]HF: heart failure.
[c]HFrEF: heart failure with reduced ejection fraction.
[d]AS: aortic stenosis.
[e]AF: atrial fibrillation.
[f]CIED: cardiac implantable electronic device.
[g]CHD: chronic heart disease.
[h]Chol: cholesterol.
[i]VHD: valvular heart disease.
[j]CVD: cardiovascular disease.
[k]FAQ: frequently asked question.
[l]CPR: cardiopulmonary resuscitation.
[m]HEART: History, ECG, Age, Risk factors, Troponin risk algorithm.
[n]MKSAP-19: Medical Knowledge Self-Assessment Program (19th edition).
[o]BLS: Basic Life Support.
[p]PALS: Pediatric Advanced Life Support.
[q]MI: myocardial infarction.
[r]PAD: peripheral arterial disease.
[s]VV: varicose veins.
[t]CP: cardiovascular prevention.

XSL•FO
RenderX

Table . Outcomes of large language model applications.

Group and manuscript	Quantitative
Chronic and progressive cardiac conditions	
Dimitriadis et al [27]	ChatGPT-3.5 answered 43/47 (91%) HF[a] patient questions adequately; 4/47 (9%) right but insufficient.
Riddell et al [33]	ChatGPT-4 responses to FAQs[b]: 71% (50/70, median FRE[c] 40.2, grade 16, IQR 48.3-34.6) at college-level readability; 23% (16/70) at recommended lower than college level and 4% (4/70) of responses requiring grade 8 - 9.
Krittanawong et al [34]	On 20 HF patient questions, ChatGPT was reliable with explanation in 40% (8/20), reliable without explanation in 40% (8/20), and unreliable in 20% (4/20).
Rouhi et al [35]	ChatGPT-3.5 simplified aortic stenosis materials to 6th-7th grade, Bard to 8th-9th; both improved from college-level baseline, all $P<.001$.
Hillmann et al [36]	On 25 AF[d] questions, ChatGPT-4 produced 84% (21/25) appropriate and 92% (23/25) comprehensible responses with 24% (6/25) missing content. On 25 CIED[e] questions, ChatGPT-4 produced 88% (22/25) appropriate and 100% (25/25) comprehensible responses with 52% (13/25) missing content.
Van Bulck and Moons [37]	40% (8/20) of experts rated ChatGPT's information as more valuable than Google, 45% (9/20) as equally valuable, and 15% (3/20) as less valuable.
Kassab et al [25]	ChatGPT-4 provided 100% (15/15) accurate responses to patient-centered questions and 73% (11/15) accurate with 27% (4/15) partly accurate responses to complex clinical scenarios, outperforming Google Bard (40% [6/15] accurate).
Han et al [24]	ChatGPT-4 achieved AUROC[f] 0.725 in the UK Biobank and 0.664 in the KoGES[g] cohort for 10-year CVD[h] risk prediction, performing comparably to the ACC[i] or AHA[j] (0.733, 0.674) and Framingham (0.728, 0.675) models.
Ali et al [23]	ChatGPT-4–assisted regression explained 34% (R^2=0.34) of the variability in age-adjusted cardiovascular mortality, with higher social vulnerability increasing mortality (β=+49.01) and greater digital literacy reducing it (β=–4.51).
Li et al [22]	On 25 cardio-oncology questions, ChatGPT-4 provided 68% (17/25) appropriate responses, followed by Bard, Claude 2, and ChatGPT-3.5 with 52% (13/25), and Llama 2 with 48% (12/25) (P=.65).
Yano et al [21]	ChatGPT-4's responses were rated appropriate in 85% (17/20) of cases, with strong interreviewer agreement (Gwet AC=0.890, SE 0.066, $P<.001$).
Kusunose et al [46]	Overall accuracy 64.5% (20/31). Accuracy was higher for clinical questions (CQs) than for limited evidence-based questions: 80% (16/20) vs 36% (4/11) (P=.005). Across 21 CQs, 9 showed zero entropy (identical answers), while 7 of the remaining 12 had entropy >0.5 (unacceptable variability).
Al Tibi et al [45]	Overall recommendations conflicted in 95% (38/40)**;** Cohen κ=−0.0127 (no agreement). Category match: stop 0%**,** decrease 0%**,** increase 6.7% (3/40)**,** add 12.5% (5/40).
Acute cardiac events	
Birkun and Gautam [38]	In 60 Bing chatbot responses, inconsistent advice appeared in 25% (5/20) of responses for the Gambia and the United States and 45% (9/20) for India. Readability required a 12th-grade level for the Gambia and the United States and 10th grade for India ($P\leq.008$).
Scquizzato et al [39]	ChatGPT-3.5 answers to cardiac arrest and CPR[k] questions were rated positively overall (mean 4.3/5, SD 0.7), with high scores for clarity (mean 4.4/5, SD 0.6), relevance (mean 4.3/5, SD 0.6), and accuracy (mean 4.0/5, SD 0.6).

XSL•FO
RenderX

Group and manuscript	Quantitative
Safranek et al [40]	ChatGPT-4 reduced nonnumerical errors from 5.7% (95% CI 3.6 - 8.9) to 0.3% (0.1 - 1.9), lowered subscore error to 0.10 (0.07 - 0.14) points with less variability (SD 0.33), and correctly classified HEART[l] risk groups in 100% (96.3 - 100) of runs, compared with 81.5% (71.7 - 88.4) for ChatGPT-3.5.
Birkun [43]	In Scenario 1, the chatbot suggested inapplicable or excessive actions in 10% (1/10) of conversations; in Scenario 2, this occurred in 30% (3/10). In Scenario 2, the chatbot failed to transition to CPR instructions after assessing the victim's condition in 30% (3/10).
Physician education	
Harskamp and De Clercq [41]	The January 2023 version of ChatGPT-3.5 performed significantly worse, answering 74% (37/50) versus 92% (46/50) of trivia questions compared with the September 2023 ChatGPT version (P=.03), and only 50% (10/20) of complex cases were answered correctly.
Skalidis et al [16]	ChatGPT answered 58.8% (213/362) correctly. ESC[m] 61.7% (42/68), BHDRA[n] 52.6% (79/150), StudyPRN 63.8% (92/144), approximating the 60% passing threshold.
Yavuz and Kahraman [15]	ChatGPT-4 received high expert agreement for differential diagnoses (median 5, IQR 1) and management plans (median 4, IQR 1), with diagnostic accuracy of 4.47, SD 0.81 in Group 1 and 4.58, SD 0.67 in Group 2, with no significant difference between groups (P<.26).
Gritti et al [13]	ChatGPT-4 answered 66% (58/88) correctly, significantly outperforming ChatGPT-3.5 at 38% (33/88); P<.001, with superior accuracy across every subspecialty topic.
Lee et al [5]	ChatGPT-4 outperformed average MKSAP-19[o] users 80% (96/120) versus 60% (72/120); P<.001; ChatGPT-3.5 also passed but lower at 55% (66/120), while PubMedGPT failed at 27% (32/120).
Patient education	
Günay et al [19]	Readability analysis showed: hospital website answers averaged a Flesch Reading Ease score of 65.6 (7th grade level), whereas ChatGPT-4 responses averaged 43.3 (11th grade level).
Bushuven et al [20]	ChatGPT-3.5 and ChatGPT-4 correctly identified the diagnosis in 94% (124/132; P=.49) of responses, but advised emergency calls in only 54% (12/22) and provided correct first aid guidance in 45% (10/22), with incorrect advanced life support instructions in 14% (3/22) of cases.
Lautrup et al [18]	ChatGPT-4 responses to 123 cardiovascular prompts averaged 3 - 4 across the 4Cs. Myocardial infarction prompts scored highest (correctness 3.84/5; conciseness 3.65/5), while cardiovascular prevention scored lowest (correctness 3.03/5; conciseness 2.71/5).
Moons and Van Bulck [17]	ChatGPT lowered readability modestly (JAMA grade 11→9; Cochrane 17→11; EJCN grade 10 unchanged) while preserving most content, with word counts changing minimally in JAMA (533→525), by 14% in Cochrane (365→315), and by 45% in EJCN (1,013→563).
Almagazzachi et al [44]	Appropriateness: 93% (93/100) overall and 7% (7/100) inappropriate, evaluated against guideline-based standards. Reproducibility: 93% (93/100) of questions reproducible and 7% (7/100) irreproducible.
Cardiac diagnostic tests	
Fijačko et al [30]	ChatGPT-4 was correct in 17/27 (63%), Bard 13/27 (48.2%), and Bing 6/27 (22.2%).
Zhu et al [42]	ChatGPT achieved 84% (21/25) overall accuracy on BLS[p] and 78.9% (30/38) on evaluable ACLS[q] items using multiple-choice inputs, improving to 96% (24/25) and 92.1% (35/38) when incorrectly answered questions were rewritten as open-ended prompts.
King et al [14]	ChatGPT-4V answered 96%(24/25) BLS and 90% (45/50) ACLS questions correctly, accuracy decreased to 75% (9/12) for questions containing ECG[r].

XSL•FO
RenderX

Group and manuscript	Quantitative
Günay et al [29]	ChatGPT-4 correctly answered 91% (36/40), outperforming emergency medicine specialists 77% (31/40, *P*<.001) and cardiologists 82% (33/40, *P*=.001).
Kangiszer et al [28]	ChatGPT-4 answered 47% (67/141) in open-ended format, 53% (75/141) in multiple choice without justification, and 55% (78/141) in multiple choice with justification formats correctly.
Sarangi et al [26]	Perplexity performed highest with 67% (50/75) concordance, followed by ChatGPT at 65% (49/75) and Bing at 63% (47/75), while Bard showed the lowest performance at 45% (34/75).

[a]HF: heart failure.

[b]FAQ: frequently asked question.

[c]FRE: Flesch Reading Ease.

[d]AF: atrial fibrillation.

[e]CIED: cardiac implantable electronic device.

[f]AUROC: area under the receiver operating characteristic curve.

[g]KoGES: Korean Genome and Epidemiology Study.

[h]CVD: cardiovascular disease.

[i]ACC: American College of Cardiology.

[j]AHA: American Heart Association.

[k]CPR: cardiopulmonary resuscitation.

[l]HEART: History, ECG, Age, Risk factors, Troponin risk algorithm.

[m]ESC: European Society of Cardiology.

[n]BHDRA: British Heart Data Research Alliance.

[o]MKSAP-19: Medical Knowledge Self-Assessment Program (19th edition).

[p]BLS: Basic Life Support.

[q]ACLS: Advanced Cardiovascular Life Support.

[r]ECG: electrocardiogram.

XSL•FO
RenderX

Table . Strengths and limitations.

Group and study	LLM[a] advantages	LLM disadvantages and limitations	Conclusion
Chronic and progressive cardiac conditions			
Dimitriadis et al [27]	Gave clear and supportive answers to common HF[b] questions.	Using a single LLM version and a fixed question source; lacked real patient interaction, no quantitative scoring.	Useful for patient education in HF, but the evidence is limited.
Riddell et al [33]	Accurate and consistent responses; objective readability assessment using validated metrics (FRE[c] and SMOG[d]).	Questions were not validated with real patients; responses were tested on a single model sample.	Reliable content, but readability should be improved for patient use.
Krittanawong et al [34]	ChatGPT model provided primarily reliable answers to commonly asked questions related to HF.	Prompts not validated with real patients; single LLM tested; lacked quantitative scoring and external validation.	Promising adjunct for HF education, not stand-alone.
Rouhi et al [35]	Used standardized readability metrics on PEMs[e] from major institutions; compared the LLMs under identical conditions.	Focused on readability only; small sample of 21 materials; limited to US sources.	Both LLMs improved the readability of aortic stenosis materials, but broader validation and comprehension testing are needed before patient application.
Hillmann et al [36]	Standardized comparison across 3 LLMs using expert-blinded evaluation.	Questions not externally validated; small dataset (50 items); limited to electrophysiology topics.	ChatGPT-4 outperformed Bing and Bard in accuracy and comprehensibility, showing strong potential for patient education.
Van Bulck and Moons [37]	ChatGPT provided clearer, more structured, and more reliable cardiology information than Google.	Very small sample (4 vignettes and 20 experts); prompts not validated with real patients.	ChatGPT is generally seen as trustworthy and useful, but evidence remains limited.
Kassab et al [25]	ChatGPT-4 is highly accurate for patient and physician queries; it outperformed Bard.	Small dataset (15 patient and 15 physician questions); subjective grading.	ChatGPT-4 is promising for patient education and clinician support in valvular disease.
Han et al [24]	Large, real-world cohorts (UK Biobank and KoGES[f]) with transparent methodology; ChatGPT-4 achieved accuracy comparable to established CVD[g] risk models.	ChatGPT-4 outputs vary with identical prompts; training data remain non transparent.	ChatGPT-4 is feasible for population-level CVD risk prediction.
Ali et al [23]	Innovative use of ChatGPT-4 for epidemiology.	Reliance on secondary data sources may introduce reporting bias; ChatGPT-4's role is limited to regression assistance.	LLMs may complement population health research, but clinical relevance is limited.
Li et al [22]	Compared multiple LLMs using standardized ESC[h] guideline–based cardio-oncology questions.	Small question set, questions researcher-generated rather than patient-derived.	ChatGPT-4 shows promise for cardio-oncology, requires oversight.
Yano et al [21]	First study to evaluate ChatGPT-4's hypertension responses in English and Japanese.	Small sample (20 questions); prompts generated by the model rather than real patients; subjective evaluation.	ChatGPT-4 provided accurate and guideline-consistent hypertension information in both languages.
Kusunose et al [46]	Potential supplementary tool for rapid access to hypertension guideline information.	Overall accuracy may be insufficient for standalone use; inconsistent answers on repeat runs (entropy); small sample; single grading evaluation; no assessment of downstream clinical outcomes.	ChatGPT may assist clinicians as a supplement, but requires caution, especially for complex questions.
Al Tibi et al [45]	Uses real-world patient data.	Single center and single cardiologist; assumes physician is correct; limited context given to ChatGPT-4; small sample.	ChatGPT-4 recommendations differed substantially from those of the cardiologist, with no agreement; further validation is needed before clinical use.
Acute cardiac events			

XSL•FO
RenderX

Group and study	LLM[a] advantages	LLM disadvantages and limitations	Conclusion
Birkum and Gautam [38]	Using repeated queries across countries.	Single nonvalidated prompt; frequent omissions of key guideline steps.	Provides relevant but often incomplete or incorrect MI[i] first aid advice, limited for unsupervised public use.
Scquizzato et al [39]	Dual evaluation by professionals and laypeople; generally positive ratings.	High reading level; subjective, unblinded scoring; single LLM version.	ChatGPT provided useful and mostly accurate CPR[j] information, but readability and safety gaps limit unsupervised use.
Safranek et al [40]	The framework prompts improvement for automated HEART[k] score determination across a limited set of synthetic patient notes.	Synthetic dataset; HEART subscores limited to structured fields.	Promising clinician decision support concept that warrants validation on real clinical data.
Birkun [43]	Natural dialogue; frequently delivered straightforward CPR steps and handled a bystander barrier by encouraging continuation.	Omitted critical elements; diagnostic risk by not asking "breathing normally"; conversational glitches and occasional inapplicable suggestions.	May be a better-than-nothing option where T-CPR[l] is unavailable, but it should not be considered reliable for real-life emergencies.
Physician education			
Harskamp and De Clercq [41]	Clear performance improvement in newer model versions.	Lower accuracy in complex consults; single run prompts; limited model transparency.	ChatGPT showed potential as AI[m] decision support for common cardiac conditions but requires further validation before clinical adoption.
Skalidis et al [16]	Large exam dataset; transparent question sourcing.	Manual single run prompting; no control of model settings.	Potential aid for exam prep.
Yavuz and Kahraman [15]	High expert agreement for differential diagnoses.	Synthetic case format may not reflect real-world nuance; modest variability in expert ratings.	Useful adjunct for training, but structured oversight is needed.
Gritti et al [13]	ChatGPT-4 markedly improved accuracy over ChatGPT-3.5; outperformed across all subspecialties; objective scoring using textbook answer key.	Moderate overall accuracy; text-only questions (no ECG or echo); single-run testing.	ChatGPT-4 performs better than ChatGPT-3.5 but remains insufficient for high-stakes pediatric cardiology use.
Lee et al [5]	Uses validated MKSAP-19 answer key.	Exam style questions only (no ECG, echo, or images); single run manual prompting.	ChatGPT-4 shows strong decision support potential, but limitations must be managed.
Patient education			
Günay et al [19]	Direct comparison with hospital websites; blinded expert rating.	Language complexity; no patient validation.	Scientifically sound but limited accessibility
Bushuven et al [20]	High diagnostic accuracy (94%) and reliable recognition of pediatric emergencies.	Simulated vignettes only; no direct comparison with humans' competence in emergency situations.	ChatGPT-4 performs better than ChatGPT-3.5 in emergency recognition but still provides incorrect and inconsistent guidance, requiring further refinement before real-world use.
Lautrup et al [18]	Innovative 4C framework; diverse prompts.	Expert-based, not real patients; no replicate testing; no control of model parameters.	Useful framework, but highlights risks of LLM-driven patient advice.
Moons and Van Bulck [17]	Clear value of LLMs for simplifying patient information and improving readability.	ChatGPT rarely reaches 6th-grade level; Bard removes large amounts of text; English-only evaluation; visuals not assessed; temperature settings not tested.	ChatGPT is useful for simplifying Patient Education Materials, but requires further evaluation.
Almagazzachi et al [44]	Large curated question set; repeated queries to assess reproducibility; dual evaluation against guidelines and physician judgment.	Predefined questions may limit topic coverage; no patient user testing; model version not specified.	ChatGPT demonstrated high accuracy and reproducibility for hypertension patient education, but human oversight remains necessary.
Cardiac diagnostic tests			

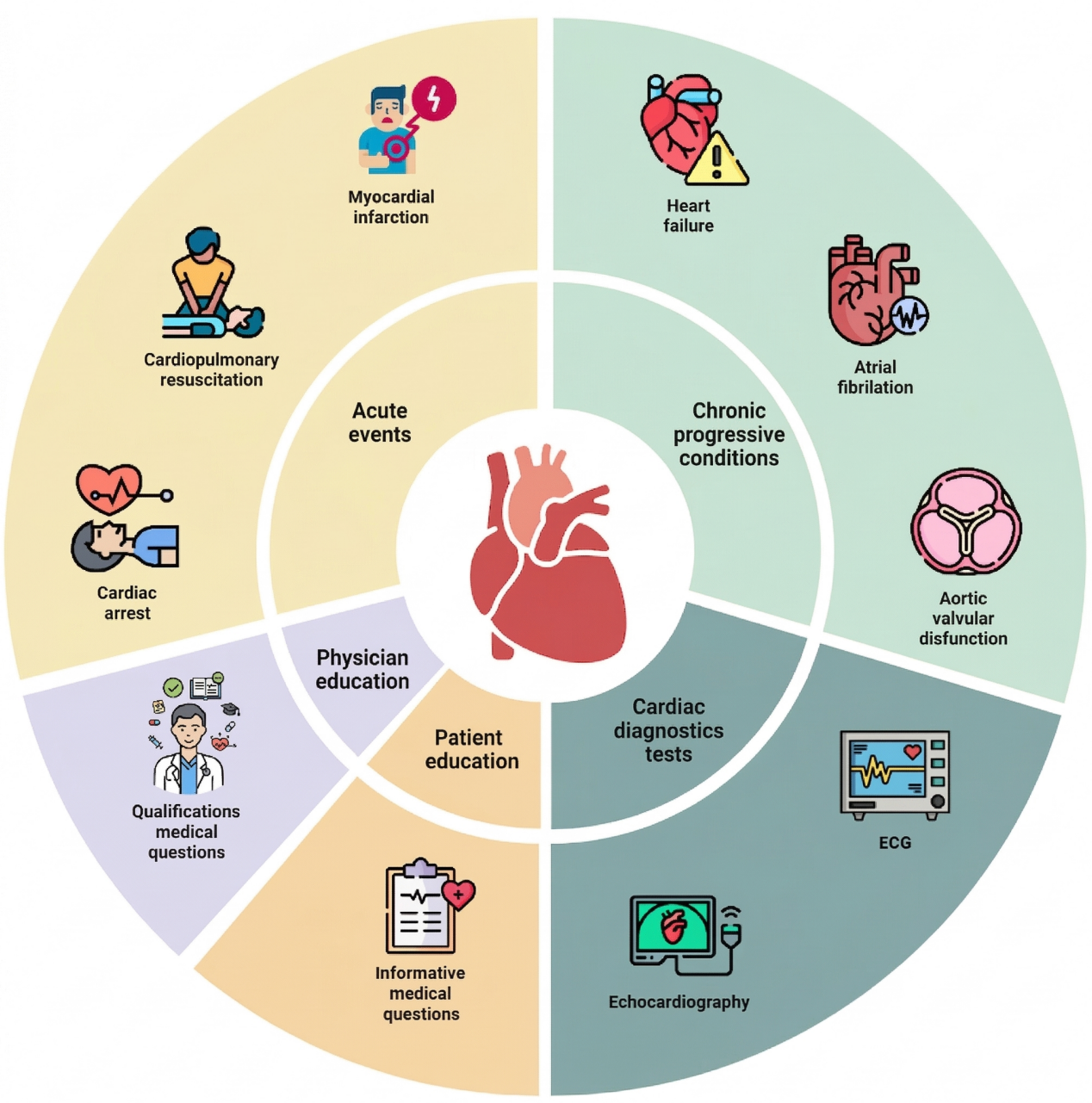
XSL•FO
RenderX

Group and study	LLM[a] advantages	LLM disadvantages and limitations	Conclusion
Fijačko et al [30]	First multimodal chatbot ECG[n] test.	Small dataset	Proof of concept multimodal LLMs can attempt ECG.
Zhu et al [42]	Open-ended prompts improved answer quality.	Open-ended prompts improved answer quality; evaluation restricted to exam-style items rather than real-world clinical variation.	ChatGPT can achieve high performance on AHA[o] exam content, especially when questions are reframed as open-ended prompts.
King et al [14]	ChatGPT-4V outperformed GPT-3.5, particularly with the inclusion of image-based questions.	No prospective testing with residents; only multiple-choice focus.	GPT-4 may support test preparation and training, but is limited to narrow tasks.
Günay et al [29]	ChatGPT showed better accuracy than the other 2 groups in everyday ECG questions.	Used text, not real ECG images; possible training exposure.	ChatGPT-4 is strong, but not a replacement.
Kangiszer et al [28]	Consistent strength in fact-based questions.	Low overall accuracy; no image interpretation capability evaluated; no comparison to human trainees.	ChatGPT-4 shows limited accuracy for echocardiography board content.
Sarangi et al [26]	Ability to generate reasonable differential diagnoses from text prompts.	Only text descriptions; 2 radiologists only.	Useful adjunct, but true concordance with experts remains limited.

[a]LLM: large language model.
[b]HF: heart failure.
[c]FRE: Flesch Reading Ease.
[d]SMOG: Simple Measure of Gobbledygook.
[e]PEM: Patient Education Materials.
[f]KoGES: Korean Genome and Epidemiology Study.
[g]CVD: cardiovascular disease.
[h]ESC: European Society of Cardiology.
[i]MI: myocardial infarction.
[j]CPR: cardiopulmonary resuscitation.
[k]HEART: History, ECG, Age, Risk factors, Troponin risk algorithm.
[l]T-CPR: telecommunicator-assisted cardiopulmonary resuscitation.
[m]AI: artificial intelligence.
[n]ECG: electrocardiogram.
[o]AHA: American Heart Association.

XSL•FO
RenderX

Figure 2. Categorization of articles into core groups with corresponding cardiology subfields.



Chronic and Progressive Cardiac Conditions

Thirteen studies [21-25,27,33-37,45,46] evaluated the application of LLMs in chronic cardiovascular disease, spanning heart failure, hypertension, valvular disease, atrial fibrillation, cardiovascular risk prediction, and cardio-oncology.

Heart failure was the most frequently studied topic. Dimitriadis et al [27] showed that ChatGPT-3.5 produced accurate answers to 91% (43/47) of patient questions, though some were incomplete. Riddell et al [33] reported that ChatGPT-4 responses to HF questions were written at a college reading level in 71% (50/70; median FRE [Flesch Reading Ease] 40.2, grade 16, IQR 48.3-34.6) of cases, while Krittanawong et al [34] reported that only 40% (8/20) of ChatGPT-3.5 responses were reliable without physician oversight.

Hypertension was addressed by Yano et al [21], who demonstrated that ChatGPT-4 produced largely appropriate answers in 85% (17/20) of hypertension-related inquiries in English and Japanese, with English responses consistently superior. Kusunose et al [46] evaluated ChatGPT-3.5 against 31 guideline-based questions derived from the Japanese Society of Hypertension (JSH) 2019 guidelines. The chatbot achieved an overall accuracy of 64.5% (20/31). Performance was significantly higher for clinical questions than for limited evidence-based questions (80% [16/20] vs 36% [4/11]; *P*=.005). A nonsignificant trend was observed for recommendation level versus evidence level questions (62% vs 38%; denominators not reported; *P*=.07). No difference was found between questions originally written in Japanese and translated questions (65% vs 58%; denominators not reported; *P*=.60). In a retrospective analysis using real-world data from a rural clinic, Al Tibi et al [45] compared antihypertensive medication

XSL•FO
RenderX

recommendations generated by ChatGPT-4 with those made by a cardiologist during laboratory review visit. Among 40 patients with hypertension, overall recommendations differed in 95% (38/40) of cases. At the level of individual medications, agreement was low, with only 10.2% of recommendations matching between ChatGPT-4 and the cardiologist (denominators not reported). The Cohen κ coefficient was −0.0127, indicating no agreement on whether to implement medication changes for a given patient.

Valvular disease was the focus of Rouhi et al [35], who showed ChatGPT-3.5 and Bard improved readability of aortic stenosis education materials, with ChatGPT-3.5 achieving the target 6th-7th grade level while Bard remained above it. Additionally, Kassab et al [25] evaluated 30 valvular disease queries, reporting ChatGPT-4 provided 100% (15/15) accurate responses to patient-centered questions and 73% (11/15) accurate and 27% (4/15) partly accurate responses to complex clinical scenarios, outperforming Google Bard (40% [6/15] accurate) and being 2.5-fold more likely to provide accurate answers (P<.001).

Atrial fibrillation and cardiac implantable device information were evaluated by Hillmann et al [36]; ChatGPT-4 produced appropriate responses in 84% (21/25) of atrial fibrillation and 88% (22/25) of cardiac implantable electronic device queries, with comprehensibility scores of 92% (23/25) and 100% (25/25), respectively. ChatGPT-4 outperformed Bing (60% [15/22], 72% [18/25] appropriate) and Bard (52% [13/25], 16% [4/25] appropriate) and showed fewer omissions and minimal confabulation.

Other chronic conditions were also explored. Han et al [24] showed ChatGPT-4 achieved 10-year cardiovascular disease risk predictions with performance similar to American College of Cardiology/American Heart Association (AHA). Li et al [22] found ChatGPT-4 outperformed other LLMs in cardio-oncology, though it was less reliable for treatment recommendations. Van Bulck and Moons [37] reported that 40% (8/20) of experts found ChatGPT's information more valuable than Google, 45% (9/20) equally valuable, and 15% (3/20) less valuable. Experts appreciated the sophistication and nuance of ChatGPT's responses but noted they were sometimes incomplete or potentially misleading.

Cardiovascular mortality was evaluated by Ali et al [23], who demonstrated the use of ChatGPT-4 to generate and execute regression models predicting mortality rates for the US county level. Across 3118 counties, the model explains 34% (R^2=0.34) of variability in age-adjusted cardiovascular mortality with higher social vulnerability increasing (β=49.01) and higher digital literacy reducing (β =−4.51) mortality.

Acute Cardiac Events

Four studies [38-40,43] evaluated LLMs in acute cardiac contexts.

First Aid in Myocardial Infarction

Birkun and Gautam [38] found that the Bing chatbot frequently omitted critical guideline concordant steps, with readability of 12th grade level for the Gambia and the United States and 10th grade for India (P≤.008). Incorrect advice appeared in 25% (5/20) of responses for the Gambia and the United States and 45% (9/20) for India.

Cardiac Arrest and Cardiopulmonary Resuscitation

Scquizzato et al [39] reported that ChatGPT-3.5 answers to cardiac arrest and cardiopulmonary resuscitation (CPR) questions were rated positively overall (mean 4.3/5, SD 0.7), with high scores for clarity (mean 4.4/5, SD 0.6), relevance (mean 4.3/5, SD 0.6), and accuracy (mean 4.0/5, SD 0.6). CPR-specific responses scored lower across all parameters, and professionals rated overall value (mean 4.0/5, SD 0.5 vs mean 4.6/5, SD 0.7; P=.02) and comprehensiveness (mean 3.9/5, SD 0.6 vs mean 4.5/5, SD 0.7; P=.02) lower than laypeople. Readability was difficult (FRE score 34 [IQR 26 - 42]). Birkun [43] assessed the New Bing chatbot's ability to provide telecommunicator-assisted CPR across 2 scenarios: Scenario 1, in which the victim was not breathing, and Scenario 2, in which the bystander was unsure whether the victim was breathing. In Scenario 2, the chatbot failed to ask for the emergency address in 50% (5/10) of cases and did not transition to CPR instructions after assessing the victim in 30% (3/10), with several additional Scenario 2 conversations reportedly interrupted or stuck at the breathing assessment step. The chatbot asked only whether the victim was "breathing" (rather than "breathing normally"), potentially missing agonal breathing and delaying arrest recognition, never inquired about nearby AED (Automated External Defibrillator) availability, and suggested inapplicable actions in 10% (1/10) of Scenario 1 and 30% (3/10) of Scenario 2.

Chest Pain Evaluation

Safranek et al [40] reported ChatGPT-4 correctly classified HEART score (History, ECG, Age, Risk factors, Troponin risk algorithm) risk groups in 100% (96.3% - 100%) of runs compared with 81.5% (71.7% - 88.4%) for ChatGPT-3.5. Iterative prompt refinement reduced nonnumerical outputs for ChatGPT-3.5 from 18.7% (95% CI 14.7 - 23.5) to 6.7% (4.4 - 10.1) and for ChatGPT-4 from 5.7% (3.6 - 8.9) to 0.3% (0.1 - 1.9).

Physician Education

Six studies [5,13,15,16,19,41] investigated the use of LLMs for supporting physician training and assessment in cardiology.

Exam preparation was assessed by Lee et al [5], who tested LLMs on 120 MKSAP-19 (Medical Knowledge Self-Assessment Program, 19th edition) cardiology questions, found that ChatGPT-4 achieved 80% (96/120), meeting the passing threshold, while PubMedGPT lagged far behind at 27% (32/120). Skalidis et al [16] reported that ChatGPT answered 58.8% (213/362) of European Exam in Core Cardiology questions correctly, close to the 60% passing threshold.

Clinical cases were addressed by Yavuz and Kahraman [15], who reported that ChatGPT-4 achieved high expert agreement for differential diagnoses (median 5, IQR 1) and management plans (median 4, IQR 1), supporting its role as a supplemental study aid, but emphasized that it should not be used unsupervised.

XSL•FO
RenderX

Clinical reasoning and decision support were tested by Harskamp and De Clercq [41]. ChatGPT-3.5 achieved correct responses in 85% (17/20) of AMSTELHEART-2 case vignettes, though performance was inconsistent in complex presentations. Gritti et al [13] found ChatGPT-4 correctly answered 66% (58/88) of pediatric cardiology cases, compared with 38% (33/88; *P*<.001) for ChatGPT-3.5, with superior accuracy across all subspecialty topics, when the passing threshold was set at 70%. Both models produced explanations containing incorrect or inconsistent reasoning, which were not formally graded.

Chest pain information was assessed by Günay et al [19], who found that ChatGPT-4 produced answers with comparable scientific adequacy, ease of understanding, and physician satisfaction to hospital websites (all 5.0 - 6.0/7; no significant differences), but at a much higher reading level—11th grade versus 7th grade for hospital materials.

Patient Education

Four studies [17-20] evaluated LLMs for patient information. General consultation with ChatGPT-4 was explored by Lautrup et al [18], where its responses to 123 cardiovascular prompts scored between 3 and 4 across the 4Cs (correctness 3.45/5, conciseness 3.19/5, comprehensiveness 3.52/5, and comprehensibility 3.72/5). Performance varied by topic, with myocardial infarction prompts scoring highest (correctness 3.84/5 and conciseness 3.65/5) and cardiovascular prevention lowest (correctness 3.03/5 and conciseness 2.71/5). Higher literacy prompts yielded better responses, while lower resource language prompts unexpectedly scored higher across all domains.

Emergency situations were studied by Bushuven et al [20], who compared ChatGPT-3.5 and ChatGPT-4 in Basic Life Support (BLS) and Pediatric Advanced Life Support (PALS) cases. While both models correctly identified the diagnosis in 94% (124/132; *P*=.49) of cases, they advised calling emergency services in only 54% (12/22), provided correct first aid guidance in 45% (10/22), and gave incorrect advanced life support instructions in 14% (3/22) of cases.

Readability of patient materials was evaluated by Moons and Van Bulck [17], who found that ChatGPT improved readability with minimal content loss (JAMA 533 to 525 words; Cochrane 365 to 315 words; EJCN 1013 to 563 words), whereas Google Bard achieved lower grade levels, but removed substantial content—shortening the texts by 61% (525 to 207 words), 34% (365 to 242 words), and 80% (1013 to 204 words), often omitting important details.

Almagazzachi et al [44] compiled a final set of 100 hypertension-related questions after physician review. Each question was asked to ChatGPT 3 times, and the majority response for each question was evaluated against established reference publications. Guideline-based assessment classified 93% (93/100) of the majority of responses as appropriate and 7% (7/100) as inappropriate. A separate clinical review by 1 board-certified internal medicine physician classified 92% (92/100) as appropriate and 8% (8/100) as inappropriate, yielding an overall accuracy of 92.5% (mean of the 2 assessments). For reproducibility per question, 93% (93/100) were reproducible and 7% (7/100) were irreproducible; across all 300 responses, 3.6% (7/300) were classified as irreproducible.

Cardiac Diagnostics Tests

Six studies examined the ability of LLMs to support diagnostic testing in cardiology [14,26,28-30,42], focusing on imaging, electrocardiography, and echocardiography.

Cardiothoracic imaging was evaluated by Sarangi et al [26], who compared 4 LLMs on 25 cardiac differential diagnosis items. ChatGPT and Perplexity provided more consistent differential diagnoses than Bing and Bard (67% [50/75] vs Bing 63% [47/75] and Bard 45% [34/75]), though accuracy remained moderate and dependent on case complexity.

Electrocardiography performance was assessed across several studies. Fijačko et al [30] evaluated multimodal LLMs on ECG image interpretation, finding that ChatGPT-4 Pro correctly interpreted 63% (17/27) of ECG images, outperforming Google Bard 48% (13/27) and Bing 22.2% (6/27). Zhu et al [42] subsequently assessed ChatGPT-4 on AHA BLS or ACLS (Advanced Cardiovascular Life Support) examination items. Although the model achieved 84% (21/25) accuracy on BLS items and 78.9% (30/38) on evaluable ACLS questions using multiple choice prompts, the majority of its errors originated from ECG-containing items. Accuracy improved substantially to 96% (24/25) for BLS and 92.1% (35/38) for ACLS, when incorrectly answered multiple-choice questions were rewritten as open-ended prompts. More recently, King et al [14] evaluated ChatGPT-4V on the full 75-item AHA BLS or ACLS examination, achieving 96% (24/25) accuracy on BLS and 90 % (45/50) on ACLS items, with performance decreasing to 75% (9/12) on questions that contained ECG strips in the ACLS examination.

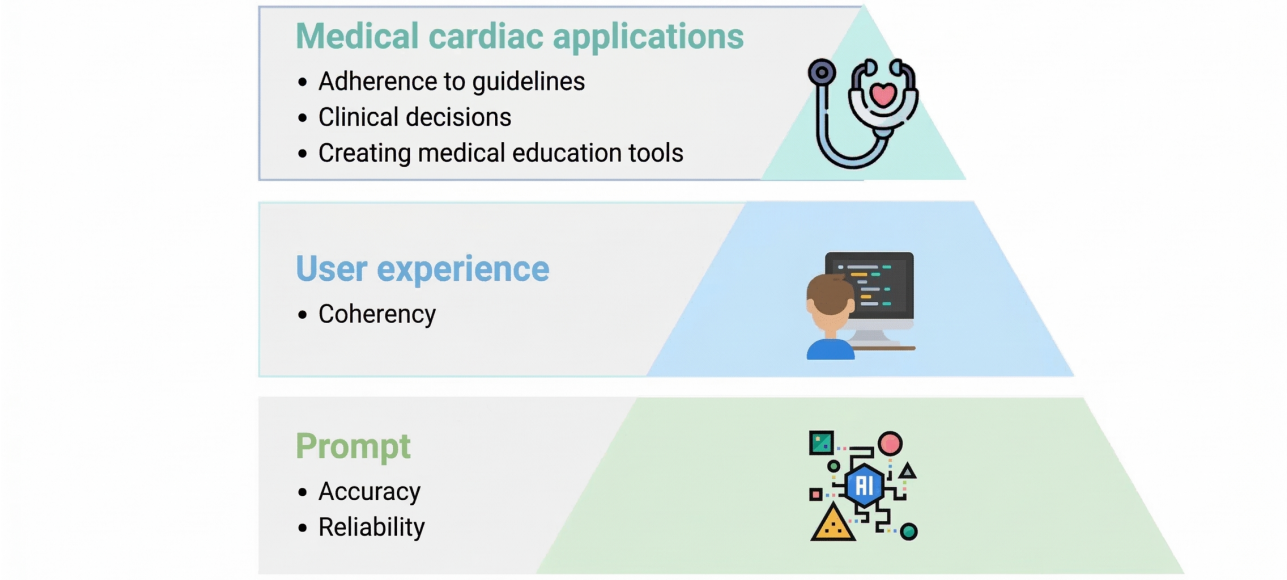
In a separate vignette-based evaluation, Günay et al [29] tested ChatGPT-4 on 40 written ECG case scenarios, finding 91% (36/40) accuracy exceeding that of emergency physicians 77% (31/40; *P*<.001) and comparable to cardiologists 82% (33/40; *P*=.001), though the model consistently struggled with wide QRS tachycardias.

Echocardiography was evaluated by Kangiszer et al [28], who tested ChatGPT-4 on 150 echocardiography board-style questions, answered 141. Accuracy remained modest, with ChatGPT-4 correctly answering 47.3% (67/141) of open-ended items, 53.3% (75/141) of multiple-choice items without justification, and 55.3% (78/141) with forced justification. Overall performance was inadequate for board-level competency.

Figure 3 outlines the 3 primary aspects explored in the articles, including the reliability of LLMs, user interaction, and their specific applications in cardiology.

XSL•FO
RenderX

Figure 3. Illustration of 3 key focus areas regarding large language models in cardiology.



Discussion

This systematic review included 35 studies [5,13-46] on the use of LLMs in cardiology, grouped into 5 domains: chronic and progressive cardiac conditions, acute cardiac events, physician education, patient education, and cardiac diagnostic tests. Overall, the studies showed that LLMs have potential across multiple aspects of cardiac management. In chronic conditions, models such as ChatGPT accurately answered common patient questions and improved the readability of educational materials [27], supporting patient engagement in long-term care. In acute emergencies, LLMs produced advice that users found clear, relevant, and accurate [38,39], suggesting possible value for lay and professional support in time-critical situations. For cardiac diagnostics, multimodal models performed well in ECG interpretation, often matching or surpassing human specialists [14], indicating potential to support clinical workflows and reduce routine workloads.

Despite these benefits, several issues must be addressed before LLMs can be used widely in cardiology. Accuracy and consistency varied significantly across models, with some producing unreliable or inconsistent interpretations [30,36], making their use in clinical settings uncertain. Readability and accessibility also remained challenging: although some LLMs improved clarity, others achieved lower reading levels only by removing essential information [33], raising concerns for patient communication across varying literacy levels. The use of LLMs also raises data privacy concerns [40], as their deployment requires strict protection of sensitive patient information.

The reviewed articles had several limitations. Approximately half the included studies evaluated ChatGPT-3.5, whose training data extended only to September 2021, limiting its ability to provide up-to-date information, an important consideration in cardiology. Most studies relied on expert evaluations rather than patient feedback, limiting insight into real-world usability. No included studies involved actual patients, raising questions about whether individuals can engage effectively with artificial intelligence–generated content. Regarding diagnostic applications, earlier studies often relied on text-based representations of multimodal data, such as written ECG or echocardiography descriptions, which may not fully capture real-world diagnostic complexity. However, recent studies have begun directly evaluating image-based analysis using multimodal models, including assessments of ECG image interpretation by ChatGPT-4–based systems, as demonstrated in a study by King et al [14]. Despite these advances, the current evidence remains limited in scope, and performance across multimodal tasks is heterogeneous, underscoring the need for larger, standardized evaluations of direct image analysis in cardiology.

This review also has its own limitations. First, the search was restricted to PubMed and Scopus, potentially missing studies in databases such as Embase or IEEE Xplore. Second, the inclusion criteria limited the review to peer-reviewed publications, excluding conference papers and preprint repositories such as arXiv and medRxiv, where important AI-related findings are often shared prior to peer review. Third, most included articles were in silico evaluations rather than prospective trials, limiting the generalizability of the findings. The heterogeneity of tasks and methods prevented meta-analysis. Fourth, due to the rapid evolution of LLMs and changing model nomenclature, our search strategy did not incorporate newer terms such as “Copilot” or broader descriptors such as “Generative AI,” which may have resulted in missing recently published studies. Finally, because technological advancements occur quickly, relevant studies and newer LLM applications may have emerged after the search was completed.

Across studies, the adapted QUADAS (by Whiting et al at the University of Bristol) assessment revealed methodological limitations specific to LLM research. Patient selection frequently presented a high risk of bias, as many studies used researcher-generated prompts or unvalidated questions without clinician or patient confirmation. In physician education studies, question banks commonly excluded media-based content such as ECGs or echocardiography clips, restricting the range of assessed cardiology skills. The index test domain was also often

XSL•FO
RenderX

high risk, as most studies used single, nonreplicated runs without reporting temperature settings or model versioning. Reference standards were generally low risk, whereas flow, timing, and data management were limited by missing metadata, insufficient prompt transparency, and lack of full output logs.

Future research should address these limitations by conducting prospective clinical trials evaluating LLMs in real-world workflows, developing standardized metrics for accuracy, readability, and safety, and exploring electronic health record integration. Studies should also expand multimodal applications, including direct analysis of ECGs and imaging.

In conclusion, LLMs demonstrate potential in cardiology, particularly in educational applications and routine diagnostics. However, performance remains inconsistent across clinical scenarios, especially in acute care, where precision is critical. With continued refinement and responsible integration, LLMs may ultimately become valuable partners in cardiovascular care and help redefine what is possible in modern medicine.

Acknowledgments

Generative artificial intelligence (ChatGPT) was used for grammar refinement and language editing during manuscript preparation. Scientific content, including study conception, data extraction, analysis, interpretation, and revisions, was generated and verified by the authors.

Funding

This work was supported in part by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.

Data Availability

All data relevant to the study are included in the article or uploaded as supplementary information.

Authors' Contributions

Conceptualization - SS, EK
Supervision - SS, EK
Investigation, Data curation, Formal analysis - MG, SS
Validation - SS, EK, BSG
Visualization - MG
Writing – original draft – MG
Funding acquisition – GNN
Writing – review & editing - MG, GNN, KS, MC-S, BSG, OE, SS, EK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study characteristics outcomes of large language model applications, literature search strategy, and QUADAS-2 risk of bias assessment.
[DOCX File, 282 KB - cardio_v10i1e76734_app1.docx]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 23 KB - cardio_v10i1e76734_app2.docx]

References

1. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544-551. [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
2. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019 Apr;28(2):73-81. [doi: 10.1080/13645706.2019.1575882] [Medline: 30810430]
3. Wen Z, Huang H. The potential for artificial intelligence in healthcare. J Commer Biotechnol 2022;27(4):217. [doi: 10.5912/jcb1327]
4. Sarraju A, Ouyang D, Itchhaporia D. The opportunities and challenges of large language models in cardiology. JACC Adv 2023 Sep;2(7):100438. [doi: 10.1016/j.jacadv.2023.100438] [Medline: 38939505]

XSL•FO
RenderX

5. Lee PC, Sharma SK, Motaganahalli S, Huang A. Evaluating the clinical decision-making ability of large language models using MKSAP-19 cardiology questions. JACC Adv 2023 Nov;2(9):100658. [doi: 10.1016/j.jacadv.2023.100658] [Medline: 38938709]
6. Ahn C. Exploring ChatGPT for information of cardiopulmonary resuscitation. Resuscitation 2023 Apr;185:109729. [doi: 10.1016/j.resuscitation.2023.109729] [Medline: 36773836]
7. Boonstra MJ, Weissenbacher D, Moore JH, Gonzalez-Hernandez G, Asselbergs FW. Artificial intelligence: revolutionizing cardiology with large language models. Eur Heart J 2024 Feb 1;45(5):332-345. [doi: 10.1093/eurheartj/ehad838] [Medline: 38170821]
8. Gala D, Makaryus AN. The utility of language models in cardiology: a narrative review of the benefits and concerns of ChatGPT-4. Int J Environ Res Public Health 2023 Jul 25;20(15):6438. [doi: 10.3390/ijerph20156438] [Medline: 37568980]
9. Sharma A, Medapalli T, Alexandrou M, Brilakis E, Prasad A. Exploring the role of ChatGPT in cardiology: a systematic review of the current literature. Cureus 2024;16(4). [doi: 10.7759/cureus.58936]
10. Moher D, Shamseer L, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4(1):1. [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
11. Schiavo JH. PROSPERO: an international register of systematic review protocols. Med Ref Serv Q 2019;38(2):171-180. [doi: 10.1080/02763869.2019.1588072] [Medline: 31173570]
12. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
13. Gritti MN, AlTurki H, Farid P, Morgan CT. Progression of an artificial intelligence chatbot (ChatGPT) for pediatric cardiology educational knowledge assessment. Pediatr Cardiol 2024 Feb;45(2):309-313. [doi: 10.1007/s00246-023-03385-6] [Medline: 38170274]
14. King RC, Bharani V, Shah K, Yeo YH, Samaan JS. GPT-4V passes the BLS and ACLS examinations: an analysis of GPT-4V's image recognition capabilities. Resuscitation 2024 Feb;195:110106. [doi: 10.1016/j.resuscitation.2023.110106] [Medline: 38160904]
15. Yavuz YE, Kahraman F. Evaluation of the prediagnosis and management of ChatGPT-4.0 in clinical cases in cardiology. Future Cardiol 2024 Mar 11;20(4):197-207. [doi: 10.1080/14796678.2024.2348898] [Medline: 39049771]
16. Skalidis I, Cagnina A, Luangphiphat W, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? Eur Heart J Digit Health 2023 May;4(3):279-281. [doi: 10.1093/ehjdh/ztad029] [Medline: 37265864]
17. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept. Eur J Cardiovasc Nurs 2024 Mar 12;23(2):122-126. [doi: 10.1093/eurjcn/zvad087] [Medline: 37603843]
18. Lautrup AD, Hyrup T, Schneider-Kamp A, Dahl M, Lindholt JS, Schneider-Kamp P. Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. Open Heart 2023 Nov;10(2):e002455. [doi: 10.1136/openhrt-2023-002455] [Medline: 37945282]
19. Günay S, Yiğit Y, Halhalli HC, Tulgar S, Alkahlout BH, Azad AM. AI in patient education: assessing the impact of ChatGPT-4 on conveying comprehensive information about chest pain. Am J Emerg Med 2024 Mar;77:220-221. [doi: 10.1016/j.ajem.2023.12.047] [Medline: 38242775]
20. Bushuven S, Bentele M, Bentele S, et al. "ChatGPT, Can You Help Me Save My Child's Life?" - diagnostic accuracy and supportive capabilities to lay rescuers by ChatGPT in prehospital basic life support and paediatric advanced life support cases - an in-silico analysis. J Med Syst 2023 Nov 21;47(1):123. [doi: 10.1007/s10916-023-02019-x] [Medline: 37987870]
21. Yano Y, Nishiyama A, Suzuki Y, et al. Relevance of ChatGPT's responses to common hypertension-related patient inquiries. Hypertension 2024 Jan;81(1):e1-e4. [doi: 10.1161/HYPERTENSIONAHA.123.22084] [Medline: 37916418]
22. Li P, Zhang X, Zhu E, et al. Potential multidisciplinary use of large language models for addressing queries in cardio-oncology. J Am Heart Assoc 2024 Mar 19;13(6):e033584. [doi: 10.1161/JAHA.123.033584] [Medline: 38497458]
23. Ali MM, Gandhi S, Sulaiman S, Jafri SH, Ali AS. Mapping the heartbeat of America with ChatGPT-4: unpacking the interplay of social vulnerability, digital literacy, and cardiovascular mortality in county residency choices. J Pers Med 2023 Nov 21;13(12):1625. [doi: 10.3390/jpm13121625] [Medline: 38138852]
24. Han C, Kim DW, Kim S, et al. Evaluation of GPT-4 for 10-year cardiovascular risk prediction: insights from the UK Biobank and KoGES data. iScience 2024 Feb 16;27(2):109022. [doi: 10.1016/j.isci.2024.109022] [Medline: 38357664]
25. Kassab J, Kapadia V, Massad C, et al. Comparative analysis of chat-based artificial intelligence models in addressing common and challenging valvular heart disease clinical scenarios. J Am Heart Assoc 2023 Nov 21;12(22):e031787. [doi: 10.1161/JAHA.123.031787] [Medline: 37982246]
26. Sarangi PK, Irodi A, Panda S, Nayak DSK, Mondal H. Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: perspectives of four large language models. Indian J Radiol Imaging 2024 Apr;34(2):269-275. [doi: 10.1055/s-0043-1777289] [Medline: 38549881]
27. Dimitriadis F, Alkagiet S, Tsigkriki L, et al. ChatGPT and patients with heart failure. Angiol Open Access 2025 Sep;76(8):796-801. [doi: 10.1177/00033197241238403]

XSL•FO
RenderX

28. Kangiszer G, Mahtani AU, Pintea M, et al. Low performance of ChatGPT on echocardiography board review questions. JACC Cardiovasc Imaging 2024 Mar;17(3):330-332. [doi: 10.1016/j.jcmg.2023.09.004] [Medline: 37943230]
29. Günay S, Öztürk A, Özerol H, Yiğit Y, Erenler AK. Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. Am J Emerg Med 2024 Jun;80:51-60. [doi: 10.1016/j.ajem.2024.03.017] [Medline: 38507847]
30. Fijačko N, Prosen G, Abella BS, Metličar Š, Štiglic G. Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images? Resuscitation 2023 Dec;193:110009. [doi: 10.1016/j.resuscitation.2023.110009] [Medline: 37884222]
31. Zhu L, Mou W, Wu K, Zhang J, Luo P. Can DALL-E 3 reliably generate 12-lead ECGs and teaching illustrations? Cureus 2024 Jan;16(1):e52748. [doi: 10.7759/cureus.52748] [Medline: 38384621]
32. Mehta S. Exploring angina: a fascinating chat with ChatGPT. Curr Probl Cardiol 2024 Mar;49(3):102393. [doi: 10.1016/j.cpcardiol.2024.102393] [Medline: 38232926]
33. Riddell CW, Chan C, McGrinder H, Earle NJ, Poppe KK, Doughty RN. College-level reading is required to understand ChatGPT's answers to lay questions relating to heart failure. Eur J Heart Fail 2023 Dec;25(12):2336-2337. [doi: 10.1002/ejhf.3083] [Medline: 37964183]
34. Krittanawong C, Rodriguez M, Kaplin S, Tang WHW. Assessing the potential of ChatGPT for patient education in the cardiology clinic. Prog Cardiovasc Dis 2023;81:109-110. [doi: 10.1016/j.pcad.2023.10.002] [Medline: 37832625]
35. Rouhi AD, Ghanem YK, Yolchieva L, et al. Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. Cardiol Ther 2024 Mar;13(1):137-147. [doi: 10.1007/s40119-023-00347-0] [Medline: 38194058]
36. Hillmann HAK, Angelini E, Karfoul N, Feickert S, Mueller-Leisse J, Duncker D. Accuracy and comprehensibility of chat-based artificial intelligence for patient information on atrial fibrillation and cardiac implantable electronic devices. Europace 2023 Dec 28;26(1):euad369. [doi: 10.1093/europace/euad369] [Medline: 38127304]
37. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. Eur J Cardiovasc Nurs 2024 Jan 12;23(1):95-98. [doi: 10.1093/eurjcn/zvad038] [Medline: 37094282]
38. Birkun AA, Gautam A. Large language model-based chatbot as a source of advice on first aid in heart attack. Curr Probl Cardiol 2024 Jan;49(1 Pt A):102048. [doi: 10.1016/j.cpcardiol.2023.102048] [Medline: 37640177]
39. Scquizzato T, Semeraro F, Swindell P, et al. Testing ChatGPT ability to answer laypeople questions about cardiac arrest and cardiopulmonary resuscitation. Resuscitation 2024 Jan;194:110077. [doi: 10.1016/j.resuscitation.2023.110077] [Medline: 38081504]
40. Safranek CW, Huang T, Wright DS, et al. Automated HEART score determination via ChatGPT: honing a framework for iterative prompt development. J Am Coll Emerg Physicians Open 2024 Apr;5(2):e13133. [doi: 10.1002/emp2.13133] [Medline: 38481520]
41. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). Acta Cardiol 2024 May;79(3):358-366. [doi: 10.1080/00015385.2024.2303528] [Medline: 38348835]
42. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation 2023 Jul;188:109783. [doi: 10.1016/j.resuscitation.2023.109783] [Medline: 37349064]
43. Birkun A. Performance of an artificial intelligence-based chatbot when acting as EMS dispatcher in a cardiac arrest scenario. Intern Emerg Med 2023 Nov;18(8):2449-2452. [doi: 10.1007/s11739-023-03399-1] [Medline: 37603142]
44. Almagazzachi A, Mustafa A, Eighaei Sedeh A, et al. Generative artificial intelligence in patient education: ChatGPT takes on hypertension questions. Cureus 2024 Feb;16(2):e53441. [doi: 10.7759/cureus.53441] [Medline: 38435177]
45. Al Tibi G, Alexander M, Miller S, Chronos N. A retrospective comparison of medication recommendations between a cardiologist and ChatGPT-4 for hypertension patients in a rural clinic. Cureus 2024 Mar;16(3):e55789. [doi: 10.7759/cureus.55789] [Medline: 38586651]
46. Kusunose K, Kashima S, Sata M. Evaluation of the accuracy of ChatGPT in answering clinical questions on the Japanese society of hypertension guidelines. Circ J 2023 Jun 23;87(7):1030-1033. [doi: 10.1253/circj.CJ-23-0308] [Medline: 37286486]

Abbreviations

ACLS: Advanced Cardiovascular Life Support
AED: Automated External Defibrillator
AHA: American Heart Association
BLS: Basic Life Support
CPR: cardiopulmonary resuscitation
ECG: electrocardiogram
FRE: Flesch Reading Ease
HEART: History, ECG, Age, Risk factors, Troponin risk algorithm

XSL•FO
RenderX

JSH: Japanese Society of Hypertension
LLM: large language model
MeSH: Medical Subject Headings
MKSAP-19: Medical Knowledge Self-Assessment Program, 19th edition
PALS: Pediatric Advanced Life Support
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Coristine; submitted 29.Apr.2025; peer-reviewed by RSG Mahmoud, SA Popoola, Y Zhang; revised version received 16.Jan.2026; accepted 19.Jan.2026; published 16.Apr.2026.

Please cite as:
Gendler M, N Nadkarni G, Sudri K, Cohen-Shelly M, S Glicksberg B, Efros O, Soffer S, Klang E
Large Language Models in Cardiology: Systematic Review
JMIR Cardio 2026;10:e76734
URL: https://cardio.jmir.org/2026/1/e76734
doi:10.2196/76734

© Moran Gendler, Girish N Nadkarni, Karin Sudri, Michal Cohen-Shelly, Benjamin S Glicksberg, Orly Efros, Shelly Soffer, Eyal Klang. Originally published in JMIR Cardio (https://cardio.jmir.org), 16.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Applications of Smart Textiles for Ambulatory Electrocardiogram Monitoring: Scoping Review of the Literature

Clarissa Pedrini Schuch[1,2], PhD; Gabriela Chaves[1], PhD; Bastien Moineau[1], PhD; Sarah Bennett[1], MSc; Meysam Pirbaglou[1], PhD; Edwin Martin Lobo[3], BSc; Milad Alizadeh-Meghrazi[1], PhD

[1]Research and Development, Myant Health Corp, 2660 Speakman Dr, Mississauga, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Software Quality Assurance, Myant Health Corp, Mississauga, ON, Canada

Corresponding Author:
Clarissa Pedrini Schuch, PhD
Research and Development, Myant Health Corp, 2660 Speakman Dr, Mississauga, ON, Canada

Abstract

Background: Smart textiles (ie, electronic textiles) offer a promising solution to ease continuous electrocardiogram (ECG) monitoring, but their real-world clinical application has been limited.

Objective: This review comprehensively examines the current state of research on textile-based ECG monitoring systems, synthesizing current evidence with respect to performance (ie, signal quality, function under static and dynamic conditions), user experience, and current challenges.

Methods: A systematic literature search across the PubMed, MEDLINE, and Embase databases from 2000 to 2025 identified 34 research papers eligible for inclusion.

Results: Textile-based ECG electrodes demonstrated good signal quality and comfort, particularly under static conditions. Nonetheless, integration into clinical practice requires addressing critical issues, which include greater efforts at validating these technologies in clinical settings and populations, as well as ensuring data security, cost-effectiveness, user-friendliness, and data interoperability.

Conclusions: Considering the prominence of feasibility research, the successful clinical integration of textile-based ECG monitoring systems requires comprehensive efforts at establishing a clinical evaluation research base (via clinical trials) and developing regulatory policies.

(JMIR Cardio 2026;10:e74261) doi:10.2196/74261

KEYWORDS

cardiac monitoring; smart textiles; electrocardiogram; ECG; remote patient monitoring; wearable; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA

Introduction

Cardiovascular disease is a leading cause of morbidity and mortality globally, necessitating the development of improved diagnostic and management tools. In recent years, driven by advances in wearable sensors, wireless connectivity, and cloud computing, there have been substantial developments in large-scale ambulatory collection, transmission, and monitoring of various physiological and clinical outcomes [1-3]. As integrated in telemedicine and remote patient monitoring solutions, these advances offer significant improvements to cardiovascular disease care, where continuous monitoring is crucial for early disease detection and timely intervention [1,2]. Among wearable technologies, electronic textile sensors have emerged as a promising solution for unobtrusive, continuous physiological monitoring in free-living environments [3,4], including in ambulatory electrocardiogram (ECG) monitoring [5,6].

ECG monitoring relies on the detection of the heart's electrical activity via surface electrodes (placed on the skin), and subsequent transmission of the signal for analysis [7]. Long-term ECG monitoring is vital for the early detection of cardiac arrhythmias, enabling the prompt treatment and prevention of severe complications. Advanced textile–based ECG systems, often termed "smart clothing" or "smart textiles," integrate sensors and electronics within wearable garments [8-10]. These systems offer the potential for continuous ECG monitoring using dry electrodes, which are embedded in conductive fabrics or textiles [11]. This approach avoids the discomfort and limitations of conventional gel (Ag or AgCl) electrodes and conductive gels [11,12], providing a convenient method for long-term monitoring [5,11,13]. Smart textiles can additionally reduce the need for invasive monitoring modalities (eg, loop

XSL•FO
RenderX

and event monitors) and provide a better diagnostic yield (due to the ease of wearing).

Despite enthusiasm, however, adoption of textile-based ECG monitoring in clinical practice remains limited. While extensive research supports the technology's technical feasibility in measuring ECG [5,6,10], few clinical studies have assessed the validity of this technology in real-world settings [14]. Challenges hindering widespread clinical adoption include the lack of rigorous clinical validation research and compliance with stringent regulatory requirements (eg, data privacy and security, data interoperability with electronic health records [EHRs]). It is important to acknowledge that the successful translation of textile-based ECG monitoring from research prototypes to widely adopted clinical and consumer applications also hinges on the development of user-friendly technologies and user acceptance. On the technical front, maintaining adequate electrode-skin contact pressure for reliable signal acquisition and ensuring the preservation of garment integrity and performance through repeated washing cycles are critical considerations. While addressing these aspects is not the primary focus of this review, technical developments represent active areas of research, closely linked with user comfort and the long-term viability of textile-based ECG systems. This review aims to map and synthesize the existing scientific literature on textile-based ECG monitoring systems, as integrated in clothing and garments.

Methods

This review aims to map the scientific literature on textile-based ECG monitoring systems integrated into clothing and garments. Research studies, published in the English language between 2000 and 2025, that investigated the use of textile-based wearable technologies for ambulatory ECG monitoring in humans were included in this scoping review. Studies were excluded if they focused solely on nontextile wearables. The search time frame (2000-2025) was chosen to capture recent advances in smart textiles, excluding outdated technologies. The language restriction was applied to ensure feasibility and maintain the consistency of data interpretation. The results were reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Checklist 1).

The search strategy used a combination of keywords and controlled vocabulary terms (MeSH and Emtree) across PubMed, MEDLINE, and EMBASE (via OVID) databases. The details of the search are described in Multimedia Appendix 1. Key terms included "Smart textile," "e-textile," "smart clothing," "textile electrode," "wearable technology," "biosensors," "cardiac rehabilitation," "remote monitoring," "ambulatory monitoring," "Holter," "cardiovascular diseases," "arrhythmia," "atrial fibrillation," "heart disease," and "cardiac disorders." Filters for human studies and publication year limits were subsequently applied to ensure adherence with the specified inclusion criteria.

The selection process involved 3 stages. First, all identified records were checked for duplicate publications. Second, following the removal of duplicate records, titles and abstracts of the remaining records were screened for relevance to the research question and the eligibility criteria. Third, full-text papers of potentially relevant studies were reviewed to confirm eligibility. Any discrepancies were resolved through discussion among the review team members. This process ensured a rigorous selection of studies appropriate for inclusion in the scoping review. In total, across 981 records identified across databases, 63 records were selected for full-text review of eligibility. Following this review, 34 studies were selected for inclusion (Figure 1).

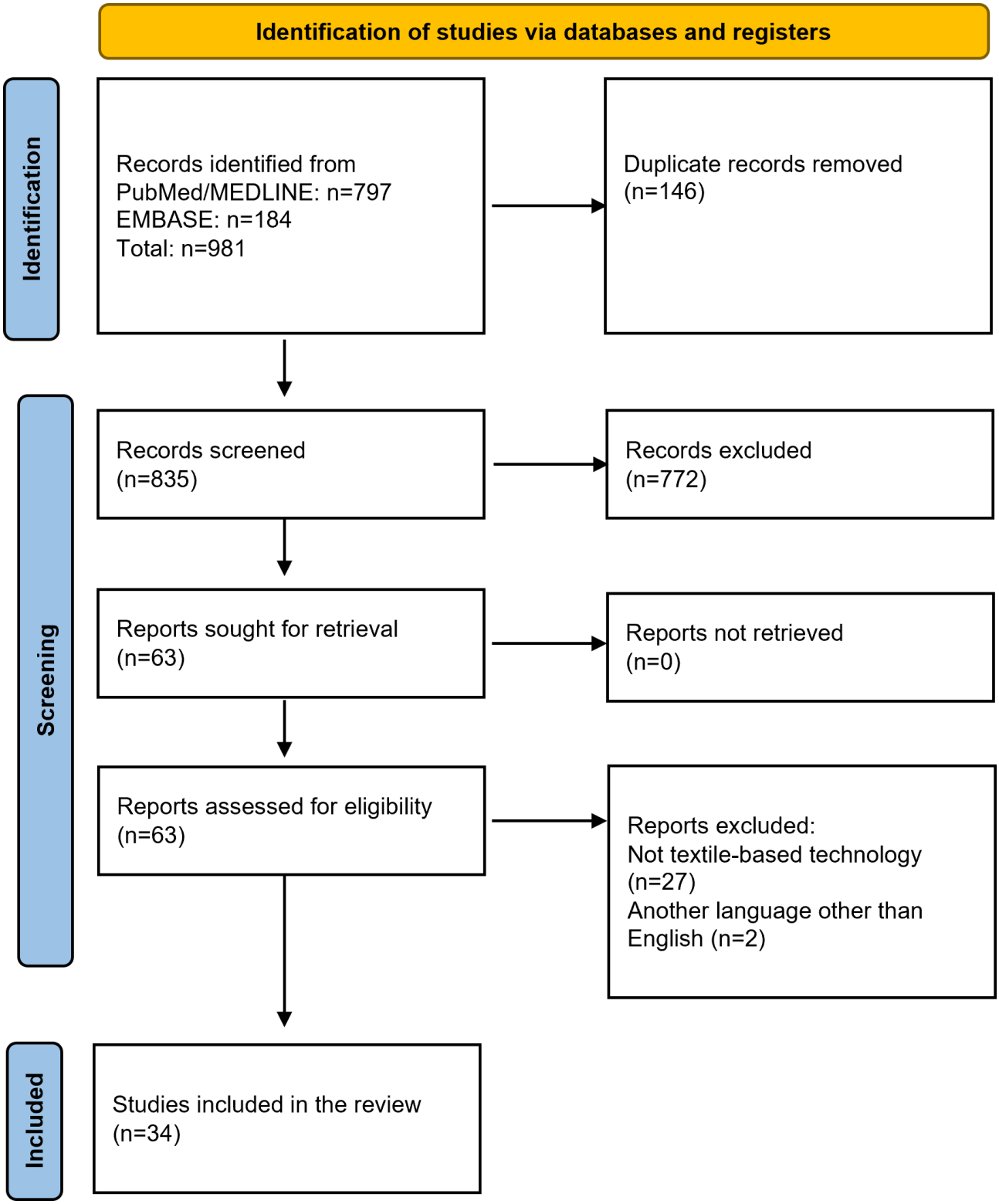
Given the limited number of published randomized clinical trials focusing on textile-based ECG in real-world clinical settings, this scoping review included feasibility studies, with a focus on ECG signal quality and user comfort as key outcome measures. This methodological choice acknowledges potential limitations in generalizability. It emphasizes the need for further research that incorporates diverse populations and real-world conditions to fully evaluate the clinical utility of this technology. The limited number of studies reporting on clinical outcomes directly reflects the current state of research in this area. The 34 studies were then comprehensively synthesized with respect to 3 key areas of research in smart textile–based ECG monitoring:

1. Comparative signal quality: To evaluate the signal quality of smart textile electrodes to traditional gel electrodes in evaluating various ECG indices (ie, QRS complex, P-wave, T-wave, R-peak).
2. Performance in static and dynamic conditions: To evaluate the performance of textile electrodes under static (eg, resting) and dynamic (eg, movement) conditions, examining the impact of motion artifacts on signal quality.
3. User experience: To evaluate user comfort, which is a critical factor for long-term adoption and use of smart textile–based ECG monitoring platforms, considering factors such as fabric type, electrode placement, and overall wearability.

Finally, this review outlines a roadmap for key clinical research priorities, together with highlighting the challenges that hinder the wider adoption and integration of textile-based ECG monitoring systems.

XSL•FO
RenderX

Figure 1. Flowower diagram of search strategy.



Results

Research Landscape

Between 2000 and 2025, smart textile technologies for ECG monitoring have seen steady progress, with research efforts predominantly based on prototype development [15-33] and observational research studies [34-44]. Smart textile technologies reviewed here use a range of conductive materials (eg, silver, copper, stainless steel, and metal inks), as incorporated into different garment types (eg, knitted patches, bands, vests, t-shirts, and bras). Among key design challenges for e-textiles is reliable electrode adhesion, particularly given the contrasting properties of metallic electrodes and the nonmetallic nature of clothing. To address this, various integration techniques are used, including direct knitting or weaving of electrodes into the fabric, embroidery of conductive yarns, adhesive bonding with specialized durable adhesives, or

XSL•FO
RenderX

printing with conductive inks requiring encapsulation. Garment design, such as the use of compression garments, and material selection, prioritizing rough electrode surfaces for improved mechanical interlocking, also play crucial roles. Different electrode materials and garment designs influence the contact pressure between the electrode and the skin. Contact pressure has an important role in ensuring the proper placement of the wearable and providing comfort. Excessive pressure can improve signal quality but can also increase discomfort and cause skin damage, while insufficient pressure can compromise the device's functionality [40]. Some studies have investigated compression garments to enhance contact pressure, while others have focused on optimizing electrode design to distribute pressure evenly. While these techniques offer promising solutions, the long-term durability of these adhesion methods and their resistance to repeated washings remains a key focus for future research. Wash durability is a critical consideration for textile-based ECG systems. Repeated washing cycles can degrade conductive materials, alter garment shape, and compromise electrode-skin contact. Research is ongoing to identify durable and washable electrode materials, as well as garment construction techniques that minimize performance degradation over time [45].

Although recent reviews, such as that by Xu et al [46], have explored the landscape of smart textiles in sports and health care, this work distinguishes itself by providing a focused analysis of textile-based systems specifically engineered for ECG monitoring, with a particular emphasis on clinical translation and utility. As evidenced by the analysis in Table 1 (static conditions) and Table 2 (dynamic conditions), the selection of electrode type significantly influences the performance and applicability of these systems. This segmentation by static and dynamic conditions is crucial because static conditions provide a controlled environment to assess basic electrode performance, while dynamic conditions, which introduce motion artifacts and other real-world challenges, are essential for evaluating the true utility of these systems. For example, silver-coated yarns are widely used in static settings, but their susceptibility to motion artifacts can be a limiting factor in dynamic environments. Novel materials, such as conductive polymers, hold promise for improved flexibility and signal stability, though further research is needed to validate their long-term performance in both static and dynamic scenarios. This highlights the importance of carefully considering the intended application and monitoring environment when selecting the optimal electrode technology for textile-based ECG monitoring.

XSL•FO
RenderX

Table . Studies summarized by static monitoring setting.

Author	Study type	Sample size	Electrode type	Findings	Condition
Mestrovic et al [26] (2007)	Observational	31	Textile-based piezoresistive transducer (vest)	Readable signals in cardiac inpatients lying	Static
Vojtech et al [30] (2013)	Prototype or concept	1	Silver-coated polyester and cotton nanoparticles	The amplitude of the individual waves of ECG signal obtained using the textile electrodes was smaller than the reference signal	Static
Dai et al [21] (2016)	Prototype or concept	6	Flexible polypyrrole textile electrodes	Heart rate accuracy of each subject was more than 95%	Static
Kannaian et al [29] (2012)	Prototype or concept	3	Embroidered the conductive yarn on polyester fabric	Electrocardiogram signal from the textile electrode was similar to the gelled electrode	Static
Pandian et al [20] (2008)	Prototype or concept	25	Sensors (silicon rubber with pure silver fillings)	RR[a] intervals presented good stability	Static
Tsukada et al [36] (2019)	Observational	66	T-shirts for men and brassieres for women with textile electrodes	P-wave, QRS, and T-wave were comparable between the textile and conventional electrodes	Static
Nigusse et al [19] (2020)	Prototype or concept	6	Silver-printed textile electrodes	The signals collected by silver-printed textile electrodes showed clear R-peaks without missing and false peaks, indicating good waveforms	Static
Arquilla et al [35] (2020)	Observational	8	Textile electrodes (silver thread) were sewn in a chest-mounted configuration	The sewn electrodes produced clean data, allowing the measurement of beat-to-beat variability and average heart rate	Static
Fouassier et al [25] (2020)	Feasibility or pilot	30	Smart t-shirt composed of 13 textiles electrodes made of silver yarns and hydrogel pads that released water vapor	A sinus rhythm was distinguished for all recordings obtained with both smart textile and reference system for all 3 static conditions	Static

[a]RR: R-R interval.

XSL•FO
RenderX

Table . Studies summarized by dynamic monitoring setting[a].

Author	Study type	Sample size	Electrode type	Findings	Condition
Hsu et al [24] (2019)	Prototype or concept	3	Noncontact dry electrodes in an elastic chest vest and attached with Velcros (not embedded)	ECG signal quality obtained by the noncontact electrode was similar to the conventional gel electrode. ECG signal quality was also good while walking.	Dynamic
Paradiso et al [15] (2005)	Prototype or concept	1	Fabric with strain sensors	Metal-based fabric electrodes reliably recorded bioelectric potentials.	Dynamic
Perez de Isla et al [39] (2011)	Feasibility or pilot	31	Noncontact dry electrodes in an elastic chest vest	Good intermethod agreement for common ECG parameters. Nubo shirt allowed pretest and posttest echocardiography. The system offers continuous, noninvasive, remote monitoring. The initial device had frequency limitations (0.5-100 Hz), affecting ST-segment analysis.	Dynamic
Trindade et al [17] (2016)	Prototype or concept	5	T-shirts comprised of 5 skin-contact textile electrodes	T-shirt prototypes provided adequate performance in standing states. Motion artifact interference, mainly caused by friction between the textile electrodes and the skin, considerably limited the performance of the prototypes in mobile contexts.	Dynamic
Pola et al [27] (2007)	Prototype or concept	4	Silver electrodes	Jogging caused the biggest problem in measurements.	Dynamic
Gonzales et al [31] (2015)	Prototype or concept	Not reported	T-shirt with 3 dry silver-based electrodes	Good efficacy of dry electrodes even in a high-motion environment.	Dynamic
Li et al [18] (2020)	Prototype or concept	1	Textile electrodes (silver-coated nylon yarns) knitted into a T-shirt	The smart clothing showed good performance for measuring ECG signals.	Dynamic
Ghahjaverstan et al [43] (2023)	Feasibility pilot	20	Dry electrodes made from conductive yarns, knitted within an elastic chest band	HR[b] measurements comparable to the reference system across all tasks, including exercising.	Dynamic

XSL•FO
RenderX

Author	Study type	Sample size	Electrode type	Findings	Condition
Romagnoli et al [47] (2014)	Observational	12	Smart textile system (GOW) with embedded textile electrodes in shirt, transmitting data to an electronic module	Excellent agreement between GOW and ECG for RR[c] intervals (LoAs[d] around ±3 ms). Good agreement for MeanRR, SDNN[e], SD2. Poor agreement for RMSSD[f], HF[g], LF/HF[h], HFnu[i], and SD1 (wide limits of agreement). The GOW system showed overall good accuracy for HR but limited accuracy for many HRV[j] parameters.	Dynamic
Coyle et al [38] (2010)	Prototype or concept	Not reported	Piezoresistive sensor	ECG signals gathered during rest showed high quality.	Dynamic
Di Rienzo et al [16,41] (2005, 2006)	Prototype or concept; observational (development phases for the same system)	1; 31	Textile-based piezoresistive transducer (vest)	Readable signals in cardiac inpatients pedaling.	Dynamic
Di Rienzo et al [48] (2013)	Observational	40	Textile vest with knitted conductive-fiber electrodes (single-lead ECG)	Readable ECG 95.9% vs 82.7% (traditional); artifact rate 4.1% vs 17.3%; arrhythmia detection sensitivity 99.7%, specificity 99.9%; PQ/QRS comparable; RR interval error ~1 ms	Dynamic
Weder et al [32] (2015)	Prototype or concept	12	Two Ag- or Ti-coated polyethylene terephthalate electrodes embedded into a chest belt	The embroidered electrodes presented signals comparable with Ag or AgCl gel electrodes	Dynamic
Pani et al [33] (2016)	Prototype or concept	10	Patch of textile electrodes made by treating conventional fabrics with a highly conductive solution of PEDOT:PSS	Textile electrodes showed similar performance, or even better, in wet conditions	Dynamic
Alizadeh Meghrazi et al [13] (2020)	Prototype or concept	6	Silver-plated nylon yarns and carbon-coated nylon yarns knitted in a waistband	ECG signals were reliably obtained from different locations on the waist	Dynamic
Fukuma et al [49] (2019)	Feasibility or pilot	100	Textile electrodes (Hitoe fabric with PEDOT-PSS) embedded in a t-shirt	The system was suitable for lifestyle or sports activities, demonstrating AF[k]-detection performance similar to that of other wearable devices	Dynamic
Pagola et al [50] (2023)	Feasibility or pilot	163	Textile electrodes in a garment	The textile wearable Holter monitoring detected pAF[l] in 35.37% of patients	Dynamic

XSL•FO
RenderX

Author	Study type	Sample size	Electrode type	Findings	Condition
Teferra et al [34] (2021)	Feasibility or pilot	1	Textile electrodes embedded in a smart ECG vest	Performance was comparable to a traditional 3-lead Holter monitor	Dynamic
Machino et al [42] (2023)	Randomized controlled trial	67	Dry textile electrodes (PEDOT-PSS and nanofiber) embedded in a garment	Garment ECG (2-week) detected AF recurrence in 18% of patients, significantly higher than 24-hour Holter (6%)	Dynamic
Amami et al [44] (2022)	Feasibility or pilot	31	Silver textile electrodes embedded in an undershirt	Wearable undershirt electrode demonstrated abilities comparable to Holter ECG for ECG monitoring and arrhythmia detection	Dynamic
Olmos et al [51] (2014)	Observational	31	nECG SHIRT: Biomedical shirt with integrated textile electrodes (BlendFix) for ECG signal acquisition	Excellent correlation between Nuubo system and conventional tilt table test for commonly assessed ECG parameters during tilt testing	Dynamic
Yu et al [52] (2017)	Cohort study	5	Wearable 12-lead ECG T-shirt with 10 dry textile electrode patches (Shieldex Medtex P180) and active electrodes	Average per-lead signal coverage ranged from 20.9% to 56.3%. After combining data from all leads (temporal fusion), overall coverage improved to up to 81.9%. A 3-stage artifact detection algorithm effectively identified artifacts from 50-Hz noise and motion.	Dynamic
Pagola et al [22] (2018)	Feasibility or pilot	146	Three-lead ECG vest and 1-lead ECG chest band	Both garments were similarly comfortable during the day and night. However, the vest group presented a longer time of compliance and time analyzed than the chest band. The percentage of missed signal was lower in the vest group. The rate of undiagnosed AF detected with textile Holter was 21.9%	Dynamic
Steinberg et al [23] (2019)	Prototype or concept	15	Bra or shirt with integrated sensors recording a single-lead ECG	The wearable ECG sensor's signal quality and accuracy were equivalent to Holter monitoring. Signal coverage of R-R intervals showed a very close overlay between the wearable sensor and Holter signals. The wearable sensor presented high wearing comfort and minimal risk of skin irritation.	Dynamic

XSL•FO
RenderX

Author	Study type	Sample size	Electrode type	Findings	Condition
Neri et al [53] (2024)	Observational	30	Crop top garment with embedded polymer–based electrodes	YouCare System showed 70% "Good," 12% "Acceptable," and 18% "Not Readable" ECG signals	Dynamic

[a]Prototype or concept: Early engineering or technical validation studies, often with small samples or bench testing. Feasibility or pilot: Preliminary human studies assessing usability, signal quality, or performance in real-world conditions. Clinical studies: Observational, cohort, or randomized controlled designs.

[b]HR: heart rate.

[c]RR: R-R interval.

[d]LoA: limits of agreement.

[e]SDNN: standard deviation of NN intervals.

[f]RMSSD: root mean square of successive differences.

[g]HF: high-frequency component.

[h]LF/HF: low-frequency/high-frequency ratio.

[i]HFnu: normalized high-frequency component.

[j]HRV: heart rate variability.

[k]AF: atrial fibrillation.

[l]pAF: paroxysmal atrial fibrillation.

Comparative Signal Quality of Textile-Based ECG Monitoring and Standard Gel Electrodes

The standard ECG signal comprises morphological features (ie, P-wave, QRS, and T-wave), reflecting atrial and ventricular depolarization and repolarization, and temporal features (ie, PR, QRS, QT, and R-R intervals), representing durations between these events [5]. These features provide key insights into cardiovascular physiology and pathology, making continuous long-term ECG monitoring essential for detecting cardiovascular disorders and abnormalities [54].

Traditionally evaluated over the long term using Holter monitors, ambulatory ECG monitoring relies on adhesive gel (Ag or AgCl) electrodes, which provide excellent signal quality [24,35]. However, long-term use is often limited by skin irritation, gel drying, and electrode detachment due to perspiration [24,35,55]. Smart textile electrodes offer a potential solution to these issues by using conductive fabrics, materials, or metals to acquire ECG signals without the need for gels [11].

The demonstration of the accuracy and reliability of smart textile–based ECG monitoring platforms is essential to enhance their adoption in clinical settings. In relation to identifying morphological and temporal ECG features, most studies have focused on the QRS complex, P-wave, T-wave, and R-peak amplitude variations and smart textiles' signal-to-noise ratio. For example, in an early demonstration of the feasibility of textile-based ECG monitoring, Di Rienzo et al [16,56] used a conductive fiber vest to accurately detect arrhythmias and heart rate compared to a traditional ECG recorder, while Paradiso et al [15] showed comparable results with a metal-yarn vest. Further research investigated various conductive materials (eg, silver-coated nylon, stainless steel yarn [26]), electrode placement (chest bands, t-shirts, wristbands [26,27,29,30]), and different textile constructions, demonstrating good signal quality in many studies in agreement with reference devices [13,19,21,23,26,29,30,35,42-44,47,50,52]. However, some studies reported issues, such as noise from dry skin [27], signal amplitude discrepancies due to electrode placement [30], and waveform irregularities [18]. While many studies focused on signal quality and accuracy, and some incorporated signal processing algorithms to improve data quality [15,16,39,57], few addressed the integration of textile-based ECG data into clinical practice [34,42-44,50,58-61].

Several studies demonstrated good textile-based ECG signal quality [20,21,26,29,32,34,36,38,46,52,57], highlighting the technology's promise for long-term monitoring. However, limitations remain: data transmission methods (eg, Bluetooth, Wi-Fi) often require extensive postprocessing; most studies were conducted in controlled laboratory settings with healthy participants and did not use fully integrated wearable form factors; and the clinical validation and long-term feasibility remain insufficiently explored. Future research should address these limitations through more comprehensive real-world studies incorporating diverse populations and integrated clinical workflows to ensure reliable and meaningful data.

Textile-Based ECG Monitoring Performance During Static and Dynamic Conditions

The division of studies based on monitoring setting (static vs dynamic) reveals key trends in the field of textile-based ECG monitoring. As demonstrated in Table 1, 9 of the reviewed studies focused on static conditions, likely reflecting the initial stages of development and validation for these technologies. The rationale for this segmentation lies in the fundamentally different challenges presented by each scenario. The static setting allows for the controlled assessment of signal quality and electrode performance, minimizing the confounding effects of motion artifacts and providing a baseline for evaluating the system's potential. However, the number of studies in the dynamic category demonstrates a growing interest in the application of these systems in real-world scenarios. This shift introduces new challenges, as the dynamic setting often leads to increased motion artifact and greater variability in

XSL•FO
RenderX

electrode-skin contact impedance. Examining the types of activities within the dynamic setting (walking, jogging, cycling, etc) further highlights these challenges, with activities involving more vigorous movement generally resulting in lower signal quality.

The inherent flexibility of textile-based ECG systems introduces challenges related to motion artifacts. Di Rienzo et al [48] explored the impact of dynamic conditions, examining the impact of motion artifacts on signal quality in telemedicine applications. The electrode-skin interface is particularly susceptible to disruption during dynamic movements, leading to noise and signal degradation, especially with dry or hairy skin [11,19,33,37]. Numerous studies have investigated the impact of both static (sitting, lying, standing) and dynamic (eg, walking, jogging, sit-to-stand) activities on ECG signal quality [17,20,24,25,31,32,34,36,38,39,41,52]. For example, Perez de Isla et al [39] evaluated a dynamic ECG monitoring system (Nuubo's dynamic ECG) against a conventional treadmill–based ECG monitoring system during exercise electrography and found comparable performance between the 2 systems on baseline and peak heart rates. Similar results were also seen by Olmos et al [51] for heart rate and maximum PR interval on the tilt table test between Nuubo's technology ECG shirt and conventional ECG monitoring.

In relation to other ECG parameters, Di Rienzo et al [48,57] also observed adequate QRS complex detection during walking and pedaling, while Pandian et al [20] reported good consistency in R-R, QRS, and QT intervals during walking with a textile vest. However, Trindade et al [17] noted higher noise amplitude with textile electrodes compared to gel electrodes during walking, and Tsukada et al [36] observed increased motion artifacts with trunk twisting. Studies included various strategies to mitigate motion artifacts, including the optimization of electrode placement, use of compression bands to enhance electrode-skin contact, and application of skin moisturizers or humidification techniques [11,17,33]. The use of a wetting device was also shown to improve signal quality in both static and dynamic conditions [32].

User Comfort

Wider clinical adoption of textile-based ECG monitoring requires user acceptance, which is linked to comfort and ease of use [11,62]. While high signal quality and accuracy are crucial, a comfortable and unobtrusive system is essential for encouraging prolonged wear and consistent data acquisition [6]. However, user comfort has been under-investigated, with only a few studies addressing this critical aspect as a secondary outcome [22,23,35,44,63].

These studies offer key insights into user perceptions. Pagola et al [22] found that patients with stroke preferred the e-textile vest over a conventional chest band for long-term use, with comparable comfort levels across day and night. Wu et al [62] reported that an e-textile t-shirt was more comfortable than traditional gel electrodes. Neri et al [53], comparing an e-textile–based ECG device to a standard Holter monitor, reported significantly greater patient comfort in the e-textile monitor, compared to the standard monitor. Steinberg et al [23] also noted greater comfort and low skin irritation with e-textile sensors, integrated into bras and shirts during 24-hour monitoring, while Arquilla et al [35] found no significant comfort difference compared to traditional electrodes. Similarly, Montazeri Ghahjaverstan et al [43] reported that all pediatric participants found the textile-based chest band easy to wear, nonirritating, and generally comfortable, with 95% rating the design as appropriate and 75% willing to use it for continuous monitoring at home.

Discussion

This review highlights the significant progress, as well as the challenges faced in optimizing e-textile–based ECG monitoring. Our findings address pertinent issues related to the optimal design and adoption of textile-based ECG monitoring technologies with respect to signal quality, static and dynamic performance, and user comfort.

Signal Quality

In relation to signal quality, the reviewed studies primarily evaluated ECG signal quality of e-textiles in healthy participants within controlled laboratory settings. The current proof-of-concept studies and feasibility testing [64,65], though essential for establishing minimum viable products, do not address the complexities of adopting e-textile–based ECG monitoring in clinical practice. Therefore, definitive conclusions about the effective use and integration of e-textiles in clinical practice cannot be drawn. Still, these studies achieved reliable ECG signals in e-textiles compared to standard adhesive electrodes and clarified key requirements to achieve reproducible ECG signals in e-textiles. These included consistent electrode placement, stable contact, supported by optimized electrode design, appropriate garment fit, garment care, and wash durability. Overall, these findings point to the need for evaluations of e-textile–based ECG monitoring in everyday life and clinical settings against standard modalities (ie, Holter monitors), incorporating diverse populations, health conditions, and standardized testing protocols.

Static and Dynamic Performance

E-textile–based ECG monitoring is particularly susceptible to motion artifacts, which disrupt ECG signals and degrade signal quality. The majority of studies included monitored ECG in static conditions, likely reflecting the initial validation efforts. While these studies indicated equal-to-superior performance for static e-textile–based ECG assessment [20,26,29,30,38,57], fewer studies included dynamic e-textile–based ECG assessment. Despite the lack of standardized protocols, many studies demonstrated the satisfactory detection of QRS complexes and rhythm abnormalities during moderate dynamic activity [20,34,38,52,57]. Nonetheless, motion artifacts significantly impacted signal quality during more strenuous activities [17,25,33,36]. A limitation of the current literature in relation to dynamic e-textile–based ECG assessment is the lack of standardized protocol, making it difficult to directly compare results across studies. Furthermore, studies in realistic, uncontrolled dynamic environments are still relatively scarce, highlighting the need for further research to address the practical challenges of implementing textile-based ECG monitoring in everyday life. In addition to factors that influence signal quality

XSL•FO
RenderX

(mentioned previously), further advancements in signal processing and machine learning are needed to reliably extract clinically meaningful data during dynamic conditions [39,60,61].

User Comfort

Although some feasibility and prototype studies report that textile electrodes may improve comfort compared to traditional gel electrodes [17,21], evidence remains limited and heterogeneous. Independent studies [22,23,35,63] have noted favorable comfort outcomes, while 1 study from our group [43] found the textile band to be "generally comfortable." These findings should be interpreted with caution given the potential for bias and the small sample sizes involved. Overall, comfort remains an underresearched area, and further independent clinical evidence is needed to confirm these preliminary observations. Future research should prioritize standardized assessment of user experience using validated tools and include diverse populations, as comfort and ease of use are critical for long-term adoption.

Clinical Integration

The potential for greater use and clinical integration of e-textiles is promising, yet significant challenges remain. These include interoperability with existing EHR systems, clinical validation, patient privacy, cost-effectiveness, timely feedback mechanisms, and user-friendliness. Effective workflow integration also requires clear standards for data review frequency, protocols for urgent findings, and sustainable reimbursement models. Frameworks, such as the ABCD model [58]—which evaluates device accuracy, clinical utility, regulatory approval, and cost—can guide adoption. Importantly, these frameworks must incorporate both clinician and patient perspectives [66-68] to ensure usability and adherence.

Recent clinical investigations provide encouraging evidence of feasibility and diagnostic value in real-world settings. Fukuma et al [49] demonstrated that a t-shirt with PEDOT-PSS electrodes was suitable for lifestyle and sports activities, achieving atrial fibrillation (AF) detection performance comparable to other wearables; although this study was included in our review, it primarily involved asymptomatic participants with cardiovascular risk factors rather than patients with established disease. Pagola et al [50] reported that intensive 90-day textile Holter monitoring detected paroxysmal AF in 35.37% of patients, underscoring the benefit of extended monitoring. Machino et al [42] showed that a 2-week garment ECG detected AF recurrence in 18% of patients—significantly outperforming 24-hour Holter monitoring (6%). Similarly, Amami et al [44] found that an undershirt with silver textile electrodes provided ECG monitoring and arrhythmia detection comparable to conventional Holter devices. Collectively, these findings indicate promising feasibility and signal quality; however, conclusions about clinical robustness cannot be drawn without formal quality assessment and larger trials.

Recommendations

The rapid evolution of telemedicine and digital health tools presents significant opportunities to improve cardiovascular disease management [69-71]. Smart textiles offer a promising pathway toward enhanced remote patient monitoring, providing benefits for patients, health care providers, and the overall health care system. However, successful integration requires addressing several key aspects.

1. Wearable data acquisition: This pillar focuses on developing robust and reliable wearable systems capable of accurately monitoring, collecting, and preprocessing ECG data from textile-based sensors. This includes considerations, such as sensor design, electrode placement, and signal quality under various conditions.
2. Data transmission and communication: The seamless and secure transmission of ECG data from the wearable device to a remote monitoring center (smartphone or PC) is crucial. This necessitates robust and reliable data communication networks, addressing issues such as signal strength, data security, and bandwidth limitations.
3. Cloud-based data analytics and clinical integration: A powerful cloud-based analytics platform is needed to process, analyze, and interpret ECG data efficiently, providing clinicians with readily accessible, accurate, and clinically relevant information [2,6,9]. This platform must facilitate timely alerts for acute events, support offline data access, and seamlessly integrate with existing EHRs. Data processing and analysis techniques are essential components of this system, requiring sophisticated algorithms to filter noise, identify key patterns, and provide meaningful interpretation [72,73]. The scalability and reliability of the backend infrastructure, including data storage and retrieval, are essential factors determining the system's performance.

Clinicians require user-friendly clinical portals to visualize and manage this data effectively, enabling timely diagnosis and intervention. Patients and caregivers should have access to relevant information for their self-management and remote health monitoring (Figure 2). Finally, user training and education are critical, particularly for individuals with low digital literacy [74]. Addressing these 3 pillars will be key to realizing the full potential of smart textiles in improving cardiovascular care.

XSL•FO
RenderX

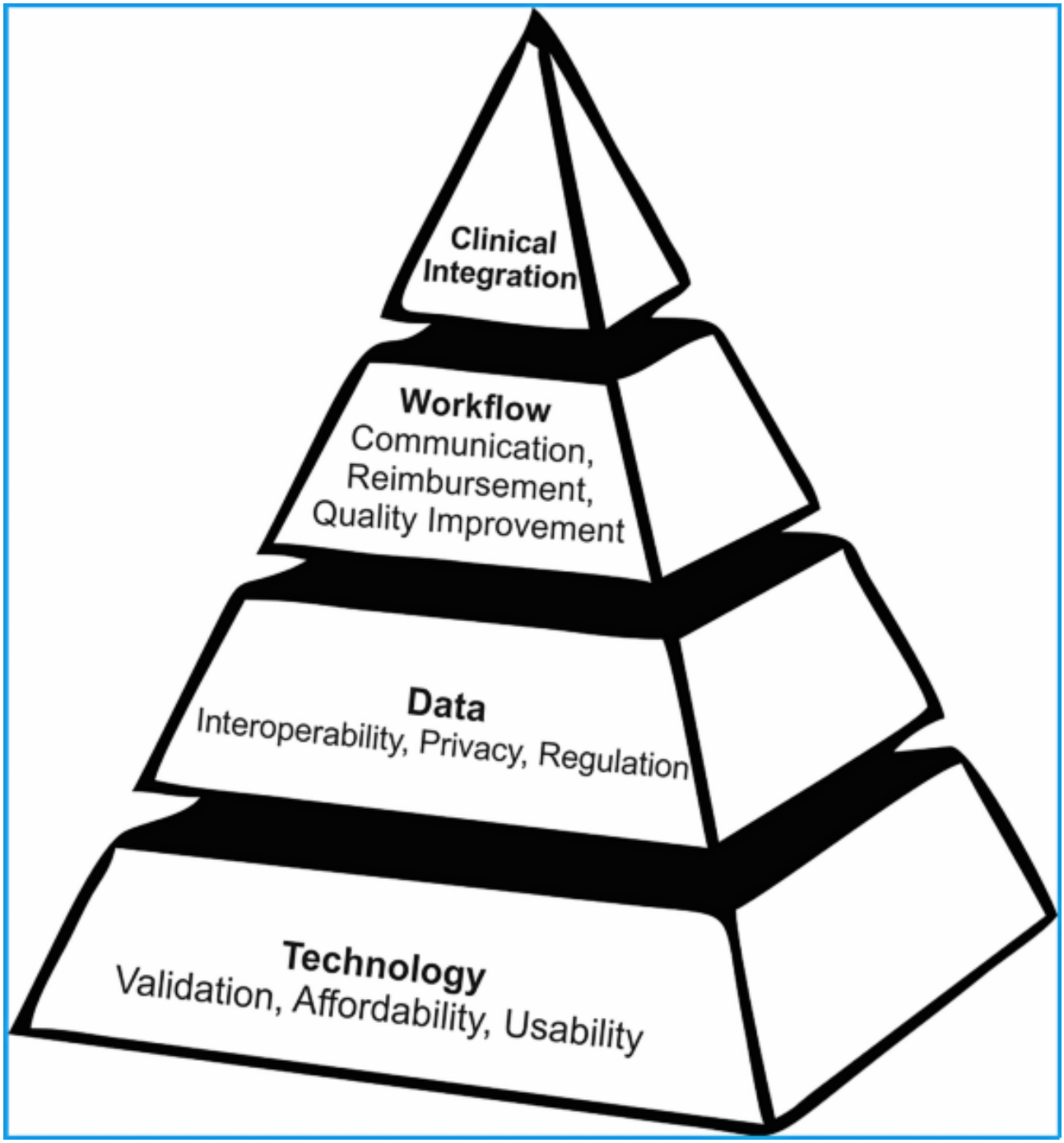
Figure 2. Smart textiles data workflow and integration in clinical practice. Textile-based sensors embedded in clothes or garments continuously measure electrocardiogram (ECG) signals.

XSL•FO
RenderX

Current Issues and Challenges

While textile-based ECG technology shows promise in terms of signal quality, several challenges must be addressed to facilitate its broader adoption in clinical practice. This section outlines key obstacles across 3 interconnected areas: technology, data management, and clinical workflow integration (Figure 3, Table 3).

Figure 3. Challenges for smart textile integration into clinical practice.



XSL•FO
RenderX

Table . Challenges and recommendations for smart textile integration into clinical practice.

Theme	Challenges for clinical integration	Recommendation
Technology	• Lack of clinical validation of smart textiles for ECG[a] monitoring. Poor data validity hampers the repeatability of measurements and impairs interoperability and data exchange. • Affordability and cost may limit the uptake of the smart textile for ECG monitoring. • Absence of information about user needs, perception, and acceptance of the smart textile for ECG monitoring.	• Extensive clinical validation regarding the safety, effectiveness, and performance of the smart textile to monitor physiological data over long periods in free-living conditions. • Private insurance companies and government may partially support the cost of smart textiles for ECG monitoring. • A comprehensive understanding of users' perceived value, perceived usefulness, perceived ease of use, and satisfaction must be met by health care professionals and users for the widespread adoption of smart textile devices. Surveys, interviews, and focus groups to address the user acceptability issue.
Data	• User privacy and security concerns related to the misuse of personal health information during data transfer and real-time data streaming. • Inefficient interoperability among devices, applications, systems, and domains. • Absence of clear regulatory policies governing smart textile for ECG monitoring.	• Clear guidelines providing the privacy, confidentiality, and proper use of electronic medical information. Ensure that users feel comfortable and confident sharing a significant amount of data with data analytic companies, health care providers, and insurance companies. • Uniform and straightforward data formats; standard communication protocols for information exchange between smart textiles and other platforms. • Adopt devices cleared or approved by regulatory agencies (i.e., FDA[b]). Expansion of HIPAA[c] policies aligned with remote patient monitoring technologies, such as smart textiles.
Clinical workflow	• Efficient processes for communicating actionable or urgent data between health care providers, patients, and caregivers are not yet well defined. In particular, the appropriate frequency of data review and the personnel responsible for monitoring incoming data still need to be established. Without clear workflows, there is a risk of data overload, which can delay timely and appropriate clinical responses. • Reimbursement structures are not fully implemented for remote patient monitoring using smart textiles technologies. • Challenges using digital health data gathered from electronic health record systems, personal health records, and other systems (eg, smart textile) for clinical quality improvement.	• Development of a clinical dashboard for data reporting. Operationalize and triage the data and define realistic response time. Health care providers can review regularly acquired ECGs during in-person or virtual visits. Timely feedback can be guaranteed if the platform allows the health care provider to set customizable notifications, for example, in case an abnormal event is detected. • Clear reimbursement models addressing remote monitoring and data review. More studies to assess the costs and cost-effectiveness of remote monitoring are necessary to determine reasonable reimbursement rates. • Clear regulations to use the smart textile data, allowing it to be aggregated, linked with other forms of data, and leveraged to generate new knowledge and promote clinical quality improvement.

[a]ECG: electrocardiogram.

[b]FDA: Food and Drug Administration.

[c]HIPAA: Health Insurance Portability and Accountability Act.

Technological Challenges

Technological challenges include the following:

- Clinical validation: While numerous studies demonstrate the feasibility of textile-based ECG electrodes in terms of signal quality, none have yet achieved routine clinical adoption (technological readiness level 9). A substantial body of clinical data is needed to validate the technology's

XSL•FO
RenderX

fidelity, interoperability, safety, and effectiveness as intended [75].

- Affordability and cost: While smart textiles could potentially reduce health care costs through optimized patient management and early intervention [61], this remains unsubstantiated. The cost-effectiveness of this technology needs further investigation, considering both initial investment and long-term operational costs [76-78]. Reimbursement policies from insurance companies and governments are crucial for broader accessibility [66,79].
- Usability and acceptability: Limited research exists on user perceptions and acceptance of smart textile ECG monitoring. Factors, such as comfort (thermal, tactile, style, pressure), ease of use, perceived value, and the ability to share data with health care providers, significantly influence user adoption [66,67,69,80,81]. Further research using user-centered design methodologies (surveys, focus groups, interviews) is essential [82,83], using frameworks, such as the technology acceptance model to understand user behavior [67,80].

Data-Related Challenges

Data-related challenges include the following:

- Privacy and security: The collection and transmission of sensitive physiological data through smart textiles raise significant privacy and security concerns [84]. Robust data protection measures, secure data storage, and transparent data usage policies are essential to safeguard patient information and ensure compliance with relevant regulations [4,58,85]. Informed consent and authorized user access are paramount.
- Interoperability: The seamless integration of smart textile ECG data with existing EHRs and other health care systems is essential for efficient remote patient monitoring [64,86,87]. This requires consistent data formats, standard communication protocols, and a cloud-based architecture that ensures data exchange and interoperability across different service providers [88].
- Regulatory policy: The rapid evolution of smart textile technology necessitates a review and update of existing regulatory frameworks [86]. Clear standards and guidelines are needed to ensure safety and effectiveness, particularly concerning data privacy and security, in accordance with regulations like Health Insurance Portability and Accountability Act and Food and Drug Administration guidelines [89]. Transparent policies regarding data collection, usage, and sharing are crucial for building trust and facilitating adoption [90].

Clinical Workflow Challenges

Clinical workflow challenges include the following:

- Communication and feedback: Effective communication and feedback between patients and health care providers are vital components of remote patient monitoring [91]. User-friendly clinical dashboards, timely alerts, and efficient reporting mechanisms are required to ensure timely intervention and improve clinical decision-making [64,73,74]. However, challenges include potential delays due to data loss, network issues, and data overload [8,70].
- Reimbursement: Clear and well-defined reimbursement pathways are essential to ensure the financial viability of smart-textile–based ECG monitoring within the health care system [74,76]. To support adoption, the cost-effectiveness of these smart wearable systems must be demonstrated, and sustainable reimbursement models need to be developed, taking into account factors such as monitoring duration and clinical service provision [92,93]. Evidence from clinician surveys also indicates support for a shared-cost approach, where expenses are distributed among patients, insurance providers, and government payers [94].
- Quality improvement: Integrating smart textile data into established clinical workflows enables continuous learning and improvement through feedback loops and data-driven insights [63,95-98]. This requires robust data analysis methods, the development of meaningful clinical indicators, and a strong focus on data-driven decision-making within the health care system. A structured approach to quality improvement, including infrastructural and organizational changes, is crucial for maximizing the benefits of smart textile technology.

Limitations

This scoping review provides a current overview of e-textile–based ECG monitoring but has several important limitations. First, our search strategy did not explicitly include ECG-specific keywords (eg, “ECG,” “EKG,” “Electrocardiogram”), which may have restricted the retrieval of engineering-focused studies emphasizing signal fidelity, such as QRS detection algorithms. While cardiovascular disease terms (eg, “Arrhythmia,” “Atrial Fibrillation”) and monitoring concepts (eg, “Holter,” “ambulatory monitoring”) were expected to capture most relevant literature, omitting explicit ECG terms could have excluded technical performance studies that do not reference disease-specific terminology. Future reviews should incorporate ECG-related keywords and engineering databases to ensure comprehensive coverage of signal processing and hardware design research.

Second, the evidence base is dominated by feasibility and proof-of-concept studies, reflecting the early developmental stage of this technology. These studies often involve small sample sizes and controlled laboratory environments, limiting generalizability to real-world clinical settings. While such controlled conditions are valuable for assessing baseline performance, they may not fully capture usability and reliability under everyday conditions. Furthermore, the absence of standardized testing protocols for key metrics, such as signal quality, wearability, and durability, makes cross-study comparisons challenging. We did not conduct a formal risk of bias assessment, so the findings should be interpreted with caution.

Third, the scope of this review was primarily focused on signal quality, performance under static and dynamic conditions, and user experience. Wash durability, a critical factor for long-term usability, was outside the main scope. Future reviews should evaluate standardized laundering protocols and durability

XSL•FO
RenderX

metrics, as these factors are essential for practical implementation and regulatory approval.

Despite these limitations, textile-based ECG monitoring demonstrates compelling potential benefits that conventional ECG methods cannot easily replicate. These include improved access to care for remote or underserved populations, continuous monitoring for the early detection of subtle cardiac changes, enhanced patient comfort and adherence due to increased wearability, and potential cost savings through earlier intervention. While further clinical validation and standardization are necessary, these advantages underscore the transformative potential of smart textiles in cardiac care.

Conclusion

In conclusion, smart textiles offer the potential for the valuable long-term monitoring of ECG parameters, improving the detection of transient events and reducing intervention times. Further, their capacity for unobtrusive, remote monitoring and seamless integration with other health metrics offers unique advantages over conventional ECG. However, widespread support and adoption within clinical workflows depend on addressing the identified limitations, incorporating user feedback, and demonstrating the clinical utility and cost-effectiveness of this technology. Only with reliable and meaningful data can textile-based ECG monitoring serve as a valuable diagnostic tool to guide care, provide a more complete picture of heart health, and improve treatment decisions.

Funding

This study was supported by the MITACS Accelerate Fellowship Program.

Conflicts of Interest

CPS is a former employee of Myant Health Corp. She also held a MITACS internship at Myant Health Corp during her postdoctoral fellowship at the University of Toronto. GC is an employee of Myant Health Corp. BM is a former employee of Myant Health Corp. MP is currently employed by Myant Health Corp. EML is a former employee of Myant Health Corp. MA-M is an employee of Myant Health Corp. SB is a former employee of Myant Health Corp.

Multimedia Appendix 1
Database search details.
[DOCX File, 20 KB - cardio_v10i1e74261_app1.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 31 KB - cardio_v10i1e74261_app2.docx]

References

1. Markert C, Sasangohar F, Mortazavi BJ, Fields S. The use of telehealth technology to support health coaching for older adults: literature review. JMIR Hum Factors 2021 Jan 29;8(1):e23796. [doi: 10.2196/23796] [Medline: 33512322]
2. Malekian Borujeni A, Fathy M, Mozayani N. A hierarchical, scalable architecture for a real-time monitoring system for an electrocardiography, using context-aware computing. J Biomed Inform 2019 Aug;96:103251. [doi: 10.1016/j.jbi.2019.103251] [Medline: 31326625]
3. Soon S, Svavarsdottir H, Downey C, Jayne DG. Wearable devices for remote vital signs monitoring in the outpatient setting: an overview of the field. BMJ Innov 2020 Apr;6(2):55-71. [doi: 10.1136/bmjinnov-2019-000354]
4. Le K, Narayana H, Servati A, et al. Electronic textiles for electrocardiogram monitoring: a review on the structure–property and performance evaluation from fiber to fabric. Textile Research Journal 2022 Feb;93(3-4):878-910. [doi: 10.1177/00405175221108208]
5. Nigusse AB, Mengistie DA, Malengier B, Tseghai GB, Langenhove LV. Wearable smart textiles for long-term electrocardiography monitoring—a review. Sensors (Basel) 2021 Jun 17;21(12):4174. [doi: 10.3390/s21124174] [Medline: 34204577]
6. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. J Am Coll Cardiol 2020 Apr 7;75(13):1582-1592. [doi: 10.1016/j.jacc.2020.01.046] [Medline: 32241375]
7. Ashley EA, Raxwal VK, Froelicher VF. The prevalence and prognostic significance of electrocardiographic abnormalities. Curr Probl Cardiol 2000 Jan;25(1):1-72. [doi: 10.1016/s0146-2806(00)70020-x] [Medline: 10705558]
8. Baig MM, Gholamhosseini H, Connolly MJ. A comprehensive survey of wearable and wireless ECG monitoring systems for older adults. Med Biol Eng Comput 2013 May;51(5):485-495. [doi: 10.1007/s11517-012-1021-6] [Medline: 23334714]
9. Duncker D, Ding WY, Etheridge S, et al. Smart wearables for cardiac monitoring—real-world use beyond atrial fibrillation. Sensors (Basel) 2021 Apr 5;21(7):2539. [doi: 10.3390/s21072539] [Medline: 33916371]

XSL•FO
RenderX

10. Stoppa M, Chiolerio A. Wearable electronics and smart textiles: a critical review. Sensors (Basel) 2014 Jul 7;14(7):11957-11992. [doi: 10.3390/s140711957] [Medline: 25004153]
11. Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):130. [doi: 10.3390/s17010130] [Medline: 28085085]
12. Yapici MK, Alkhidir TE. Intelligent medical garments with graphene-functionalized smart-cloth ECG sensors. Sensors (Basel) 2017 Apr 16;17(4):875. [doi: 10.3390/s17040875] [Medline: 28420158]
13. Alizadeh Meghrazi M, Tian Y, Mahnam A, et al. Multichannel ECG recording from waist using textile sensors. Biomed Eng Online 2020 Jun 16;19(1):48. [doi: 10.1186/s12938-020-00788-x] [Medline: 32546233]
14. Teixeira E, Fonseca H, Diniz-Sousa F, et al. Wearable devices for physical activity and healthcare monitoring in elderly people: a critical review. Geriatrics (Basel) 2021 Apr 7;6(2):38. [doi: 10.3390/geriatrics6020038] [Medline: 33917104]
15. Paradiso R, Loriga G, Taccini N. A wearable health care system based on knitted integrated sensors. IEEE Trans Inform Technol Biomed 2005 Sep;9(3):337-344. [doi: 10.1109/TITB.2005.854512]
16. Di Rienzo M, Rizzo F, Parati G, Brambilla G, Ferratini M, Castiglioni P. MagIC system: a new textile-based wearable device for biological signal monitoring. Applicability in daily life and clinical setting. Conf Proc IEEE Eng Med Biol Soc 2005;2005:7167-7169. [doi: 10.1109/IEMBS.2005.1616161] [Medline: 17281930]
17. Trindade IG, Machado da Silva J, Miguel R, et al. Design and evaluation of novel textile wearable systems for the surveillance of vital signals. Sensors (Basel) 2016 Sep 24;16(10):1573. [doi: 10.3390/s16101573] [Medline: 27669263]
18. Li M, Xiong W, Li Y. Wearable measurement of ECG signals based on smart clothing. Int J Telemed Appl 2020;2020:6329360. [doi: 10.1155/2020/6329360] [Medline: 32395127]
19. Nigusse AB, Malengier B, Mengistie DA, Tseghai GB, Van Langenhove L. Development of washable silver printed textile electrodes for long-term ECG monitoring. Sensors (Basel) 2020 Oct 31;20(21):6233. [doi: 10.3390/s20216233] [Medline: 33142899]
20. Pandian PS, Mohanavelu K, Safeer KP, et al. Smart Vest: wearable multi-parameter remote physiological monitoring system. Med Eng Phys 2008 May;30(4):466-477. [doi: 10.1016/j.medengphy.2007.05.014] [Medline: 17869159]
21. Dai M, Xiao X, Chen X, Lin H, Wu W, Chen S. A low-power and miniaturized electrocardiograph data collection system with smart textile electrodes for monitoring of cardiac function. Australas Phys Eng Sci Med 2016 Dec;39(4):1029-1040. [doi: 10.1007/s13246-016-0483-5] [Medline: 27743384]
22. Pagola J, Juega J, Francisco-Pascual J, et al. Yield of atrial fibrillation detection with textile wearable Holter from the acute phase of stroke: pilot study of Crypto-AF registry. Int J Cardiol 2018 Jan 15;251:45-50. [doi: 10.1016/j.ijcard.2017.10.063] [Medline: 29107360]
23. Steinberg C, Philippon F, Sanchez M, et al. A novel wearable device for continuous ambulatory ECG recording: proof of concept and assessment of signal quality. Biosensors (Basel) 2019 Jan 21;9(1):1-13. [doi: 10.3390/bios9010017] [Medline: 30669678]
24. Hsu CC, Lin BS, He KY, Lin BS. Design of a wearable 12-lead noncontact electrocardiogram monitoring system. Sensors (Basel) 2019 Mar 28;19(7):1509. [doi: 10.3390/s19071509] [Medline: 30925752]
25. Fouassier D, Roy X, Blanchard A, Hulot JS. Assessment of signal quality measured with a smart 12-lead ECG acquisition T-shirt. Ann Noninvasive Electrocardiol 2020 Jan;25(1):e12682. [doi: 10.1111/anec.12682] [Medline: 31339208]
26. Mestrovic MA, Helmer RJN, Kyratzis L, Kumar D. Preliminary study of dry knitted fabric electrodes for physiological monitoring. Presented at: 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information; Dec 3-6, 2007. [doi: 10.1109/ISSNIP.2007.4496911]
27. Pola T, Vanhala J. Textile electrodes in ECG measurement. Presented at: 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information; Dec 3-6, 2007. [doi: 10.1109/ISSNIP.2007.4496917]
28. Despang HG, Netz S, Heinig A, Holland HJ, Fischer WJ. Wireless long-term ECG integrated into clothing. Biomed Tech (Berl) 2008 Dec;53(6):270-278. [doi: 10.1515/BMT.2008.043] [Medline: 19037869]
29. Kannaian T, Neelaveni R, Thilagavathi G. Design and development of embroidered textile electrodes for continuous measurement of electrocardiogram signals. J Ind Text 2012 Jan;42(3):303-318. [doi: 10.1177/1528083712438069]
30. Vojtech L, Bortel R, Neruda M, Kozak M. Wearable textile electrodes for ECG measurement. Adv Electr Electron Eng 2013;11(5):410-414. [doi: 10.15598/aeee.v11i5.889]
31. Gonzales L, Walker K, Keller K, et al. Textile sensor system for electrocardiogram monitoring. 2015 Presented at: 2015 IEEE Virtual Conference on Applications of Commercial Sensors (VCACS); Mar 15 to Oct 15, 2015. [doi: 10.1109/VCACS.2015.7439568]
32. Weder M, Hegemann D, Amberg M, et al. Embroidered electrode with silver/titanium coating for long-term ECG monitoring. Sensors (Basel) 2015 Jan 15;15(1):1750-1759. [doi: 10.3390/s150101750] [Medline: 25599424]
33. Pani D, Dessi A, Saenz-Cogollo JF, Barabino G, Fraboni B, Bonfiglio A. Fully textile, PEDOT: PSS based electrodes for wearable ECG monitoring systems. IEEE Trans Biomed Eng 2016 Mar;63(3):540-549. [doi: 10.1109/TBME.2015.2465936] [Medline: 26259215]
34. Teferra MN, Hobbs DA, Clark RA, Reynolds KJ. Preliminary analysis of a wireless and wearable electronic-textile EASI-based electrocardiogram. Front Cardiovasc Med 2021;8:806726. [doi: 10.3389/fcvm.2021.806726] [Medline: 34988133]

XSL•FO
RenderX

35. Arquilla K, Webb AK, Anderson AP. Textile electrocardiogram (ECG) electrodes for wearable health monitoring. Sensors (Basel) 2020 Feb 13;20(4):1013. [doi: 10.3390/s20041013] [Medline: 32069937]
36. Tsukada YT, Tokita M, Murata H, et al. Validation of wearable textile electrodes for ECG monitoring. Heart Vessels 2019 Jul;34(7):1203-1211. [doi: 10.1007/s00380-019-01347-8] [Medline: 30680493]
37. Balsam P, Lodziński P, Tymińska A, et al. Study design and rationale for biomedical shirt-based electrocardiography monitoring in relevant clinical situations: ECG-shirt study. Cardiol J 2018;25(1):52-59. [doi: 10.5603/CJ.a2017.0102] [Medline: 28840587]
38. Coyle S, Lau KT, Moyna N, et al. BIOTEX: biosensing textiles for personalised healthcare management. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):364-370. [doi: 10.1109/TITB.2009.2038484] [Medline: 20064761]
39. Perez de Isla L, Lennie V, Quezada M, et al. New generation dynamic, wireless and remote cardiac monitorization platform: a feasibility study. Int J Cardiol 2011 Nov 17;153(1):83-85. [doi: 10.1016/j.ijcard.2011.08.074] [Medline: 21925751]
40. Alcala-Medel J, Michaelson D, Eike RJ, Li Y. Durability study of e-textile electrodes for human body communication. Text Res J 2024 Jan;94(1-2):246-258. [doi: 10.1177/00405175231197651]
41. Rienzo Di M, Rizzo F, Meriggi P, et al. Applications of a textile-based wearable system for vital signs monitoring. Presented at: Conference Proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 30 to Sep 3, 2006. [doi: 10.1109/IEMBS.2006.260173]
42. Machino T, Aonuma K, Maruo K, et al. Randomized crossover trial of 2-week Garment electrocardiogram with dry textile electrode to reveal instances of post-ablation recurrence of atrial fibrillation underdiagnosed during 24-hour Holter monitoring. PLoS ONE 2023;18(2):e0281818. [doi: 10.1371/journal.pone.0281818] [Medline: 36827294]
43. Montazeri Ghahjaverstan N, Balmer-Minnes D, Taghibeyglou B, et al. Textile-based wearable to monitor heart activity in paediatric population: a pilot study. CJC Pediatr Congenit Heart Dis 2023 Aug;2(4):187-195. [doi: 10.1016/j.cjcpc.2023.05.007] [Medline: 37969855]
44. Amami K, Yoshihisa A, Horikoshi Y, et al. Utility of a novel wearable electrode embedded in an undershirt for electrocardiogram monitoring and detection of arrhythmias. PLoS One 2022;17(8):e0273541. [doi: 10.1371/journal.pone.0273541] [Medline: 35998187]
45. Kim H, Lee S. Evaluation of laundering durability of electro-conductive textile dip-coated on para aramid knit with graphene/waterborne polyurethane composite. Fibers Polym 2018 Nov;19(11):2351-2358. [doi: 10.1007/s12221-018-8591-3]
46. Xu Z, Zhang C, Wang F, et al. Smart textiles for personalized sports and healthcare. Nanomicro Lett 2025 Apr 25;17(1):232. [doi: 10.1007/s40820-025-01749-6] [Medline: 40278986]
47. Romagnoli M, Alis R, Guillen J, Basterra J, Villacastin JP, Guillen S. A novel device based on smart textile to control heart's activity during exercise. Australas Phys Eng Sci Med 2014 Jun;37(2):377-384. [doi: 10.1007/s13246-014-0271-z] [Medline: 24756693]
48. Di Rienzo M, Racca V, Rizzo F, et al. Evaluation of a textile-based wearable system for the electrocardiogram monitoring in cardiac patients. Europace 2013 Apr;15(4):607-612. [doi: 10.1093/europace/eus368] [Medline: 23258818]
49. Fukuma N, Hasumi E, Fujiu K, et al. Feasibility of a T-shirt-type wearable electrocardiography monitor for detection of covert atrial fibrillation in young healthy adults. Sci Rep 2019 Aug 13;9(1):11768. [doi: 10.1038/s41598-019-48267-1] [Medline: 31409855]
50. Pagola J, Juega J, Francisco-Pascual J, et al. Intensive 90-day textile wearable Holter monitoring: an alternative to detect paroxysmal atrial fibrillation in selected patients with cryptogenic stroke. Heart Vessels 2023 Jan;38(1):114-121. [doi: 10.1007/s00380-022-02141-9] [Medline: 35882656]
51. Olmos C, Franco E, Suárez-Barrientos A, et al. Wearable wireless remote monitoring system: an alternative for prolonged electrocardiographic monitoring. Int J Cardiol 2014 Mar 1;172(1):e43-e44. [doi: 10.1016/j.ijcard.2013.12.068] [Medline: 24412464]
52. Yu X, Boehm A, Neu W, et al. A wearable 12-lead ECG t-shirt with textile electrodes for unobtrusive long-term monitoring—evaluation of an ongoing clinical trial. 2018 Presented at: EMBEC & NBC 2017; Jun 11-15, 2017; Tampere, Finland p. 703-706. [doi: 10.1007/978-981-10-5122-7_176]
53. Neri L, Corazza I, Oberdier MT, et al. Comparison between a single-lead ECG garment device and a Holter monitor: a signal quality assessment. J Med Syst 2024 May 27;48(1):57. [doi: 10.1007/s10916-024-02077-9] [Medline: 38801649]
54. Jacob Rodrigues M, Postolache O, Cercas F. Physiological and behavior monitoring systems for smart healthcare environments: a review. Sensors (Basel) 2020 Apr 12;20(8):2186. [doi: 10.3390/s20082186] [Medline: 32290639]
55. Marozas V, Petrenas A, Daukantas S, Lukosevicius A. A comparison of conductive textile-based and silver/silver chloride gel electrodes in exercise electrocardiogram recordings. J Electrocardiol 2011;44(2):189-194. [doi: 10.1016/j.jelectrocard.2010.12.004] [Medline: 21353065]
56. Di Rienzo M, Meriggi P, Rizzo F, et al. Textile technology for the vital signs monitoring in telemedicine and extreme environments. IEEE Trans Inf Technol Biomed 2010 May;14(3):711-717. [doi: 10.1109/TITB.2010.2048921] [Medline: 20421189]
57. Di Rienzo M, Rizzo F, Meriggi P, et al. Applications of a textile-based wearable system for vital signs monitoring. Conf Proc IEEE Eng Med Biol Soc 2006;2006:2223-2226. [doi: 10.1109/IEMBS.2006.260173] [Medline: 17946505]

XSL•FO
RenderX

58. Bayoumy K, Gaber M, Elshafeey A, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug;18(8):581-599. [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
59. Chatterjee S, Thakur RS, Yadav RN, Gupta L, Raghuvanshi DK. Review of noise removal techniques in ECG signals. IET Signal Process 2020 Dec;14(9):569-590. [doi: 10.1049/iet-spr.2020.0104]
60. Smital L, Haider CR, Vitek M, et al. Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. IEEE Trans Biomed Eng 2020 Oct;67(10):2721-2734. [doi: 10.1109/TBME.2020.2969719] [Medline: 31995473]
61. Serhani MA, T El Kassabi H, Ismail H, Nujum Navaz A. ECG monitoring systems: review, architecture, processes, and key challenges. Sensors (Basel) 2020 Mar 24;20(6):1796. [doi: 10.3390/s20061796] [Medline: 32213969]
62. Wu W, Pirbhulal S, Sangaiah AK, Mukhopadhyay SC, Li G. Optimization of signal quality over comfortability of textile electrodes for ECG monitoring in fog computing based medical applications. Future Gener Comput Syst 2018 Sep;86:515-526. [doi: 10.1016/j.future.2018.04.024]
63. Nwaru BI, Friedman C, Halamka J, Sheikh A. Can learning health systems help organisations deliver personalised care? BMC Med 2017 Oct 2;15(1):177. [doi: 10.1186/s12916-017-0935-0] [Medline: 28965492]
64. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YFY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018;1:23. [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
65. Shah SS, Gvozdanovic A. Digital health; what do we mean by clinical validation? Expert Rev Med Devices 2021 Dec;18(sup1):5-8. [doi: 10.1080/17434440.2021.2012447] [Medline: 34845960]
66. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Ind Manag Data Syst 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/IMDS-03-2015-0087]
67. Abu-Dalbouh HM. A questionnaire approach based on the technology acceptance model for mobile tracking on patient progress applications. J Comput Sci 2013 Jun 1;9(6):763-770. [doi: 10.3844/jcssp.2013.763.770]
68. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar 1;36(1):157-178. [doi: 10.2307/41410412]
69. Molinari G, Molinari M, Di Biase M, Brunetti ND. Telecardiology and its settings of application: an update. J Telemed Telecare 2018 Jun;24(5):373-381. [doi: 10.1177/1357633X16689432] [Medline: 28084886]
70. Slotwiner DJ, Tarakji KG, Al-Khatib SM, et al. Transparent sharing of digital health data: a call to action. Heart Rhythm 2019 Sep;16(9):e95-e106. [doi: 10.1016/j.hrthm.2019.04.042] [Medline: 31077802]
71. Yang Z, Zhou Q, Lei L, Zheng K, Xiang W. An IoT-cloud based wearable ECG monitoring system for smart healthcare. J Med Syst 2016 Dec;40(12):286. [doi: 10.1007/s10916-016-0644-9] [Medline: 27796840]
72. Hill DD, Garcia HM. The integration of patient health record generated from wearable and Internet of Things devices into health information exchanges. Int J Health Med Eng 2021;15:112-116 [FREE Full text]
73. Kildea J, Battista J, Cabral B, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371. [doi: 10.2196/11371] [Medline: 30741643]
74. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3(1):14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
75. Pustiek M, Beristain A, Kos A. Challenges in wearable devices based pervasive wellbeing monitoring. Presented at: 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI); Oct 22-23, 2015; Beijing, China p. 236-243. [doi: 10.1109/IIKI.2015.58]
76. Chan M, Estève D, Fourniols JY, Escriba C, Campo E. Smart wearable systems: current status and future challenges. Artif Intell Med 2012 Nov;56(3):137-156. [doi: 10.1016/j.artmed.2012.09.003] [Medline: 23122689]
77. Karunathilake SP, Ganegoda GU. Secondary prevention of cardiovascular diseases and application of technology for early diagnosis. Biomed Res Int 2018;2018:5767864. [doi: 10.1155/2018/5767864] [Medline: 29854766]
78. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21. [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
79. Ferguson C, Inglis SC, Breen PP, et al. Clinician perspectives on the design and application of wearable cardiac technologies for older adults: qualitative study. JMIR Aging 2020 Jun 18;3(1):e17299. [doi: 10.2196/17299] [Medline: 32554377]
80. Bove LA. Increasing patient engagement through the use of wearable technology. J Nurse Pract 2019 Sep;15(8):535-539. [doi: 10.1016/j.nurpra.2019.03.018]
81. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687. [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
82. Moineau B, Myers M, Ali SS, Popovic MR, Hitzig SL. End-user and clinician perspectives on the viability of wearable functional electrical stimulation garments after stroke and spinal cord injury. Disabil Rehabil Assist Technol 2021 Apr;16(3):241-250. [doi: 10.1080/17483107.2019.1668974] [Medline: 31592679]
83. Torous JB, Chan SR, Gipson SYMT, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 1;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
84. Kotz D, Gunter CA, Kumar S, Weiner JP. Privacy and security in mobile health: a research agenda. Computer (Long Beach Calif) 2016 Jun;49(6):22-30. [doi: 10.1109/MC.2016.185] [Medline: 28344359]

XSL•FO
RenderX

85. Ndlovu K, Mars M, Scott RE. Interoperability frameworks linking mHealth applications to electronic record systems. BMC Health Serv Res 2021 May 13;21(1):459. [doi: 10.1186/s12913-021-06473-6] [Medline: 33985495]
86. Herzeca LF. E-textiles: regulating the future of fashion. : Law360; 2016 URL: https://www.gibsondunn.com/wp-content/uploads/documents/publications/Herzeca-E-Textiles-Regulating-The-Future-Of-Fashion-Law360-8-25-16.pdf [accessed 2026-02-07]
87. Accenture Federal Services. Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024 draft white paper for a PGHD policy framework. 2016 URL: https://healthit.gov/wp-content/uploads/2026/01/onc_pghd_practical_guide.pdf [accessed 2026-02-17]
88. Bergmo TS. Economic evaluation in telemedicine - still room for improvement. J Telemed Telecare 2010;16(5):229-231. [doi: 10.1258/jtt.2010.009008] [Medline: 20501629]
89. Cortez A, Hsii P, Mitchell E, Riehl V, Smith P. Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024. : Office of the National Coordinator for Health Information Technology (ONC); 2018 URL: https://healthit.gov/wp-content/uploads/2026/01/onc_pghd_practical_guide.pdf [accessed 2026-02-07]
90. Connecting health and care for the nation: a shared nationwide interoperability roadmap. : The Office of the National Coordinator (ONC) for Health Information Technology; 2015 URL: https://ncvhs.hhs.gov/wp-content/uploads/2015/10/Day-2-NCVHS-Dept-Update-POSNACK.pdf [accessed 2026-02-07]
91. Elm JJ, Daeschler M, Bataille L, et al. Feasibility and utility of a clinician dashboard from wearable and mobile application Parkinson's disease data. NPJ Digit Med 2019;2(1):95. [doi: 10.1038/s41746-019-0169-y] [Medline: 31583283]
92. Physician fee schedule. Centers for Medicare & Medicaid Services (CMS). 2021. URL: https://www.cms.gov/medicare/payment/fee-schedules/physician [accessed 2026-02-07]
93. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45. [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
94. Manninger M, Zweiker D, Svennberg E, et al. Current perspectives on wearable rhythm recordings for clinical decision-making: the wEHRAbles 2 survey. Europace 2021 Jul 18;23(7):1106-1113. [doi: 10.1093/europace/euab064] [Medline: 33822029]
95. Menear M, Blanchette MA, Demers-Payette O, Roy D. A framework for value-creating learning health systems. Health Res Policy Syst 2019 Aug 9;17(1):79. [doi: 10.1186/s12961-019-0477-3] [Medline: 31399114]
96. Mason AR, Barton AJ. The emergence of a learning healthcare system. Clin Nurse Spec 2013;27(1):7-9. [doi: 10.1097/NUR.0b013e3182776dcb] [Medline: 23222020]
97. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010 Nov 10;2(57):57cm29. [doi: 10.1126/scitranslmed.3001456] [Medline: 21068440]
98. Friedman CP, Rubin JC, Sullivan KJ. Toward an information infrastructure for global health improvement. Yearb Med Inform 2017 Aug;26(1):16-23. [doi: 10.15265/IY-2017-004] [Medline: 28480469]

Abbreviations

AF: atrial fibrillation
ECG: electrocardiogram
EHR: electronic health record
MeSH : Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Coristine; submitted 20.Mar.2025; peer-reviewed by A Mills, RA Surmenev, S Mohanadas; accepted 19.Jan.2026; published 02.Mar.2026.

Please cite as:
Pedrini Schuch C, Chaves G, Moineau B, Bennett S, Pirbaglou M, Lobo EM, Alizadeh-Meghrazi M
Applications of Smart Textiles for Ambulatory Electrocardiogram Monitoring: Scoping Review of the Literature
JMIR Cardio 2026;10:e74261
URL: https://cardio.jmir.org/2026/1/e74261
doi:10.2196/74261

© Clarissa Pedrini Schuch, Gabriela Chaves, Bastien Moineau, Sarah Bennett, Meysam Pirbaglou, Edwin Martin Lobo, Milad Alizadeh-Meghrazi. Originally published in JMIR Cardio (https://cardio.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio,

XSL•FO
RenderX

is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Technologies, Clinical Applications, and Implementation Barriers of Digital Twins in Precision Cardiology: Systematic Review

Fatemeh Sarani Rad[1*], MSc; Ehsan Bitaraf[2*], MD; Maryam Jafarpour[3*], PhD; Juan Li[1*], PhD

[1]Computer Science Department, North Dakota State University, 1320 Albrecht Blvd, Fargo, ND, United States

[2]Center for Technology and Innovation in Cardiovascular Informatics, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran

[3]Center for Medical Data Science, Medical University of Vienna, Vienna, Austria

[*]all authors contributed equally

Corresponding Author:
Juan Li, PhD
Computer Science Department, North Dakota State University, 1320 Albrecht Blvd, Fargo, ND, United States

Abstract

Background: Digital twin systems are emerging as promising tools in precision cardiology, enabling dynamic, patient-specific simulations to support diagnosis, risk assessment, and treatment planning. However, the current landscape of cardiovascular digital twin development, validation, and implementation remains fragmented, with substantial variability in modeling approaches, data use, and reporting practices.

Objective: This systematic review aims to synthesize the current state of cardiovascular digital twin research by addressing 11 research questions spanning modeling technologies, data infrastructure, clinical applications, clinical impact, implementation barriers, and ethical considerations.

Methods: We systematically searched 5 databases (PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar) and screened 330 records. Forty-two original studies met the predefined eligibility criteria and were included. Data extraction was guided by 11 thematic research questions. Mechanistic and artificial intelligence (AI) or machine learning (ML) modeling strategies, data modalities, visualization formats, clinical use cases, reported impacts, limitations, and ethical or legal issues were coded and summarized. Risk of bias was evaluated using a custom checklist for modeling studies, the Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction models, and the Risk of Bias in Non-Randomized Studies - of Interventions (ROBINS-I) for observational studies.

Results: Most digital twins (29/42, 69%) relied on mechanistic models, while hybrid mechanistic–data-driven approaches and purely data-driven designs were less frequent (13/42, 31%). Only 18 studies explicitly described ML or AI, most often deep learning, Bayesian methods, or optimization algorithms. Personalization depended primarily on imaging (32/42, 76%) and electrocardiography or other electrical signals (18/42, 43%). Visualization was dominated (41/42, 98%) by static figures and anatomical snapshots. Clinically, digital twins were most commonly applied to therapy planning, risk prediction, and monitoring. Reported benefits focused on improved decision-making and therapy-related impacts, with occasional (8/42, 19%) reports of increased accuracy or faster diagnosis, but there was limited evidence for downstream improvements in patient outcomes. Key barriers included strong model assumptions and simplifications; high computational cost; data quality and availability constraints; limited external validation; and challenges in real-time performance, workflow integration, and usability. Explicit discussion of ethical, legal, or governance issues was rare (7/42, 17%).

Conclusions: Cardiovascular digital twins show substantial potential to advance precision cardiology by linking personalized physiological models with clinical decision support, particularly for therapy planning and risk prediction in arrhythmia and heart failure. However, real-world implementation is constrained by methodological heterogeneity, restricted data and validation practices, limited openness of code and models, and sparse engagement with ethical and governance questions. Future research should prioritize standardized evaluation frameworks, robust clinical validation, interoperable and user-centered system design, and ethically grounded, patient-centered development to realize the full clinical value of digital twin systems.

(JMIR Cardio 2026;10:e78499) doi:10.2196/78499

KEYWORDS

digital twin; cardiology; personalized medicine; simulation; machine learning; clinical decision support; ethics

XSL•FO
RenderX

Introduction

In recent years, the integration of digital twin technology into health care has opened new avenues for precision medicine, particularly within the field of cardiology. A digital twin is a dynamic, virtual representation of a physical system that is continuously updated with real-time data, advanced computational models, and artificial intelligence (AI) analytics [1,2]. In the context of health care, digital twins serve as virtual replicas of patients, organs, or biological systems, encompassing multidimensional, patient-specific information to inform clinical decisions [3-5].

Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality worldwide, underscoring the need for innovative, patient-centric approaches to diagnosis, treatment, and management [6,7]. The application of digital twins in cardiology involves the creation of virtual replicas of the human heart by integrating anatomical, physiological, and molecular data. These models are capable of simulating electrical activity [8], mechanical function, hemodynamics, and drug responses [9,10]. By combining data from cardiac imaging (eg, magnetic resonance imaging [MRI] and computed tomography [CT]), electrocardiography (ECG), hemodynamic profiles, electrophysiology recordings, electronic health records, and omics assessments, digital twin systems provide a basis for precision simulation and virtual experimentation [11].

These capabilities make cardiac digital twins uniquely positioned to support personalized treatment plans, enabling applications such as risk stratification, therapy optimization, surgical simulation, and drug safety testing. The integration of AI, particularly machine learning (ML) and deep learning (DL), has further improved the scalability and performance of digital twins in real-world applications.

However, despite promising technical progress, substantial challenges remain. These include (1) high computational costs and complex personalization pipelines; (2) data heterogeneity and interoperability limitations; (3) lack of standardized validation protocols and clinical benchmarks; and (4) ongoing concerns regarding privacy, explainability, and regulatory oversight.

While multiple reviews have surveyed digital twins in general health care [12] or addressed cardiovascular simulation from a technical standpoint [11], a comprehensive, domain-specific synthesis integrating technical, clinical, and implementation perspectives in personalized cardiology remains lacking.

To address this gap, we conducted a systematic review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [13,14]. This review explicitly examines original research articles on cardiovascular digital twin systems, emphasizing personalization, clinical relevance, and implementation feasibility. Our study followed a two-stage methodology:

1. Screening phase: We screened 330 articles from 5 databases (PubMed, Scopus, IEEE, Web of Science, and Google Scholar). After removing duplicates, non-English entries, and publications lacking abstracts or relevant context, 42 articles were retained.
2. Review phase: Three independent reviewers assessed full-text articles based on structured research questions (RQs). Each article was evaluated for relevance to 11 themes covering technology, data integration, clinical application, validation, ethics, and data sources.

The following RQs guided our review:

- RQ1-RQ4: What are the technological foundations of cardiovascular digital twins, including modeling strategies, AI integration, and open-source availability?
- RQ5 and RQ6: How is patient-specific data structured and visualized?
- RQ7 and RQ8: What are the clinical applications and disease targets of digital twins?
- RQ9: What clinical impacts have been reported as a result of digital twin use?
- RQ10 and RQ11: What barriers, limitations, and ethical or legal considerations are acknowledged in current studies?

The aim of this study was to systematically review the existing literature on cardiovascular digital twins to identify current technologies, clinical uses, and challenges to implementation.

Methods

Overview

This systematic review was designed and conducted following the PRISMA 2020 statement (Checklist 1). The protocol was developed in advance and used a transparent, reproducible approach to article retrieval, screening, and extraction. It was structured around 11 domain-specific RQs targeting the technological, clinical, and implementation dimensions of digital twin systems in cardiology.

Data Sources and Search Strategy

A comprehensive literature search was performed across 5 major academic databases: PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar. These platforms were selected to ensure broad interdisciplinary coverage across biomedical, engineering, and computational sciences. The databases were searched between January and early February 2025, restricting records to publications from 2010 onwards. Only the first 115 results sorted by relevance were screened for Google Scholar due to indexing limitations. The reference lists of relevant reviews were also scanned to ensure inclusion of key foundational articles.

Data collection and initial preprocessing were streamlined using Triple-A software [15], which served as the main tool for managing and organizing the retrieved records.

The search strategy used Boolean combinations of controlled vocabulary (eg, MeSH) and free-text terms as follows: (“digital twin” OR “virtual heart” OR “patient-specific model”) AND (“cardiology” OR “cardiac” OR “heart” OR “cardiovascular”) AND (“simulation” OR “personalized medicine” OR “precision medicine” OR “in silico”).

XSL•FO
RenderX

To increase transparency, we conceptually structured the search according to the Population, Intervention, Comparison, and Outcome (PICO) framework:

- Population (P): Patients with CVDs, including arrhythmia, heart failure, ischemic heart disease, cardiomyopathy, and related conditions.
- Intervention (I): Digital twin systems designed for diagnosis, simulation, personalization, monitoring, risk prediction, or therapy planning in cardiology.
- Comparison (C): Not applicable, as the review did not evaluate digital twins against alternative interventions or standard care.
- Outcome (O): Descriptive outcomes related to modeling strategies, data infrastructure, clinical applications, reported clinical impact, implementation barriers, and ethical or governance considerations.

These PICO elements informed the design of our search and eligibility criteria, while the detailed content of the review was organized around 11 thematic RQs (RQ1-RQ11).

All search results were exported to a centralized reference manager and screened using Microsoft Excel. The complete search strings for the databases are provided in Multimedia Appendix 1.

Eligibility Criteria

Articles were included if they (1) were original empirical research studies, including journal articles, conference papers, and preprints; (2) reported on the development, implementation, or evaluation of digital twin systems in health care; (3) focused on cardiovascular applications, including anatomical, physiological, or functional heart modeling; (4) were related to individualized or personalized medicine, clinical decision-making, or patient-specific therapies; and (5) were published in English and provided a structured abstract.

Articles were excluded if they (1) were review papers, commentaries, editorials, book chapters, or theoretical position pieces; (2) did not focus on cardiovascular systems (eg, neurological or orthopedic digital twins); (3) were not available in full text or lacked an identifiable abstract; (4) were duplicate entries across databases; and (5) were published in languages other than English, including those labeled as “unspecified” or “null.”

These criteria were iteratively refined during the pilot screening of 50 records.

We did not apply the exclusion criteria based on study design, as the aim of this review was to comprehensively synthesize diverse contributions to the digital twin field, including conceptual, technical, and applied studies, without limiting the scope to any particular methodological framework.

Screening and Article Selection

The initial search returned 330 records. A multistep screening protocol was applied:

- Phase 1 (title and abstract screening): Three reviewers independently screened articles for relevance. Discrepancies were resolved by group discussion and majority vote.
- Phase 2 (eligibility review): Of 44 records identified in phase 1, 2 records were excluded. A final set of 42 articles was included in the synthesis.

Reviewers used a shared Microsoft Excel spreadsheet with predefined drop-down fields for coding decisions. Interreviewer consistency was monitored, and a senior reviewer adjudicated disagreements. The filtering questions used during study selection are presented in Table 1. The complete list of all screened records, along with their inclusion or exclusion status, is provided in Multimedia Appendix 2.

Table . Filtering questions used during study selection for the systematic review of cardiovascular digital twin research.

Screening question[a]	Decision criteria
Filtering question 1: Does the study relate to digital twins in health care or medicine?	Include if the study discusses digital twins applied in health care contexts.
Filtering question 2: Does the study specifically address the use of digital twins in cardiology?	Include if the study focuses on cardiovascular applications of digital twins.
Filtering question 3: Does the study involve personalized or patient-specific applications in cardiology?	Include if the study discusses patient-specific or precision medicine approaches.

[a]Each question aligns with predefined inclusion and exclusion criteria applied across titles, abstracts, and full texts.

RQs and Data Extraction

Data extraction was organized around 11 RQs, which were structured into six thematic categories:

1. Technological foundations: modeling methods (RQ1), mechanistic model types (RQ2), ML algorithms (RQ3), and open-source availability (RQ4)
2. Data infrastructure and visualization: patient-specific data (RQ5) and visualization formats (RQ6)
3. Clinical applications and conditions: clinical use cases (RQ7) and cardiovascular conditions addressed in digital twin studies (RQ8)
4. Clinical impact: reported outcomes and benefits (RQ9)
5. Implementation challenges: technical and validation barriers (RQ10)
6. Ethical considerations: legal, privacy, and governance issues (RQ11)

Each reviewer used a structured extraction form, built in Excel, to code articles across multiple predefined categories (eg, “FEM,” “ECG,” and “Heart Failure”) using a controlled

XSL•FO
RenderX

vocabulary. Note fields allowed for contextual elaboration and inductive theme discovery.

Categories were not mutually exclusive, allowing multiple responses per article. The full data extraction form is provided in Multimedia Appendix 3.

Data Extraction Process

Data extraction followed a structured workflow as follows:

1. Full-text review: Each selected study was fully reviewed to extract methodological details and research contributions.
2. Thematic classification: Studies were assigned to predefined thematic categories based on their focus area and objectives.
3. Double-reviewer validation: Three independent reviewers extracted data; any conflicts were resolved via discussion.
4. Database compilation: Extracted data were compiled into a structured dataset for further analysis.

Risk of Bias

The risk of bias of the included studies was assessed using the instrument most appropriate for the underlying study design. Three distinct tools were used. First, simulation-based and modeling-oriented studies, such as those involving digital twins, mechanistic models, or computational pipelines, were evaluated using a custom modeling checklist, which was developed to capture methodological risks specific to computational modeling (eg, data representativeness, validation strategy, overfitting, and reproducibility). Second, prediction-modeling studies were appraised using the Prediction Model Risk of Bias Assessment Tool (PROBAST), which evaluates risk of bias across 4 domains: participants, predictors, outcome, and analysis. Finally, observational cohort studies were assessed using the Risk of Bias in Non-Randomized Studies - of Interventions (ROBINS-I), which provides structured domain-level judgments for 7 bias domains, including confounding, selection of participants, classification of interventions, missing data, and outcome measurement.

For all tools, domain-level judgments were assigned according to published guidance or tool-specific documentation. Risk-of-bias assessments were conducted independently by multiple assessors, and any discrepancies were resolved through discussion, with arbitration applied when consensus could not be reached. Domain-level ratings were then synthesized into an overall judgment (low, unclear, or high risk of bias) based on the decision rules recommended for each tool.

Visualization of risk-of-bias judgments was performed using robvis [16], an R package and web application that supports structured display of traffic-light plots and summary plots.

Results

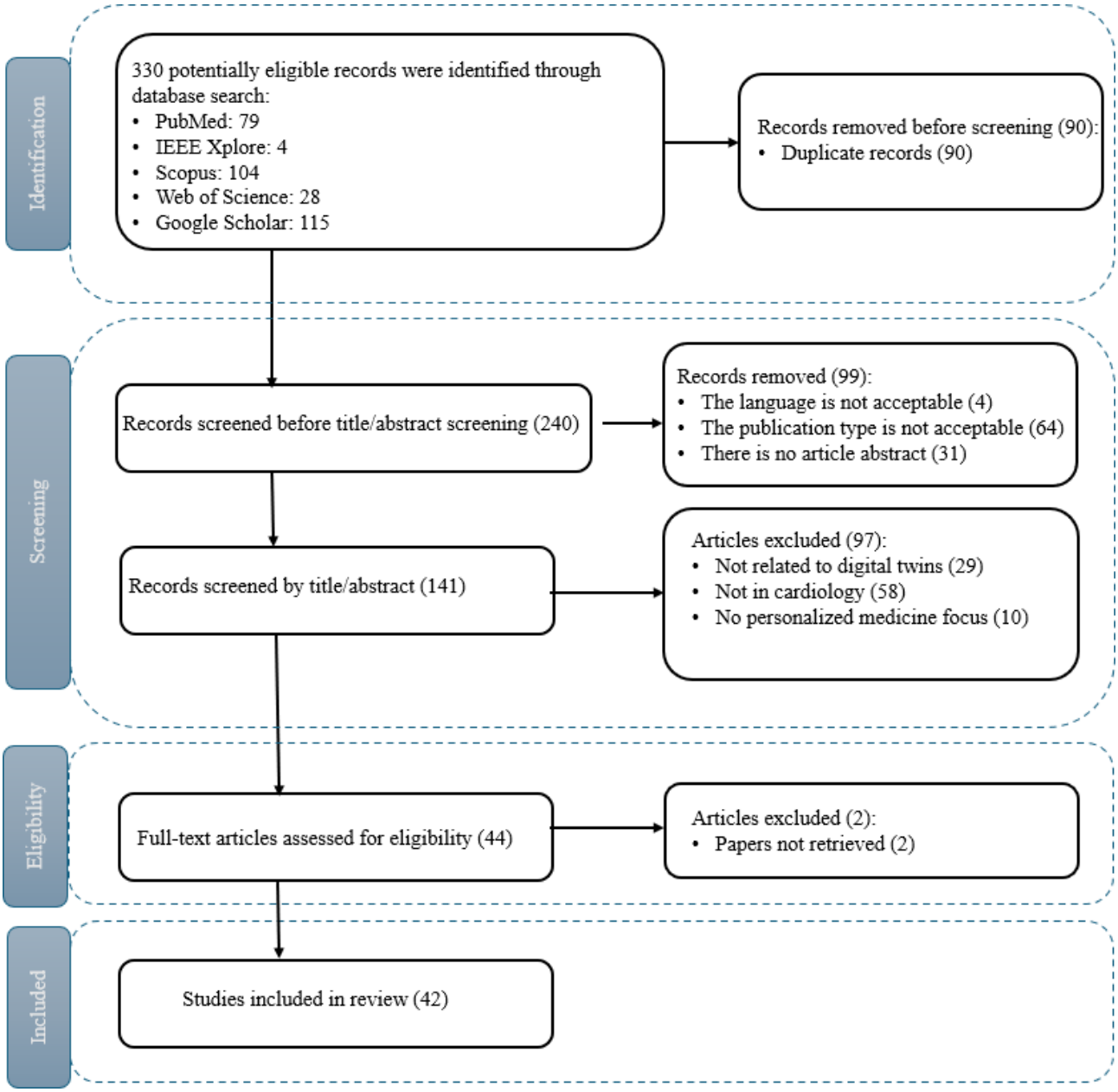
Overview

We synthesized the findings from 42 original research articles on cardiovascular digital twins. The PRISMA flow diagram of the study selection process is presented in Figure 1. The results were structured around 11 predefined RQs, which were organized into 6 thematic domains: technological foundations, data infrastructure and visualization, clinical applications and conditions, clinical impact, implementation challenges, and ethical considerations. Each subsection follows a format: overview, key insights, and interpretation. For each RQ, we present summary patterns and cite representative studies in the main text. The complete mapping of all studies to the corresponding RQ categories is provided in Multimedia Appendix 4, and the mapping from raw extraction values to the harmonized categories used in the analyses is provided in Multimedia Appendix 5.

Funding sources were reported for a subset of studies and were most often public or academic, with a smaller number supported by mixed public-foundation or public-industry collaborations and relatively few funded solely by industry. Study-level funding details are summarized in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram illustrating the systematic selection process for cardiovascular digital twin studies. A total of 330 records were retrieved from 5 major databases and screened using predefined eligibility criteria. After removal of duplicates and exclusion of irrelevant or nonoriginal articles, 42 studies were included in the final synthesis for qualitative and quantitative analyses.



Technological Foundations (RQ1-RQ4)

We outline the core technical elements of cardiovascular digital twin systems, focusing on modeling strategies (RQ1), types of mechanistic models (RQ2), ML applications (RQ3), and open-source availability (RQ4). Together, these RQs characterized how digital twins were constructed, personalized, and shared, revealing trends in hybrid modeling, the integration of AI, and the challenges in reproducibility.

RQ1: What Primary Modeling Approach is Used to Develop Digital Twins?

Overview

All 42 studies were classified according to their dominant modeling approach: mechanistic, hybrid, or data-driven. These categories reflect the computational core of digital twins, ranging from physics-based simulation to statistical learning and their integration.

Key Insights

The key insights are as follows:

- Mechanistic models were the most common (29 studies [8,11,17-43]), and they relied on physics-based formulations (eg, finite element modeling [FEM], electrophysiological simulation, and hemodynamic flow analysis) to generate personalized physiological representations.
- Data-driven models were noted in 7 studies [44-50], and they were primarily based on statistical learning or machine-learning approaches without explicit biophysical constraints.

XSL•FO
RenderX

- Hybrid approaches were the least common (6 studies [51-56]), and they combined mechanistic frameworks with data-driven components, for example, using ML to estimate parameters, extract imaging features, or accelerate computational solvers.

Interpretation

The predominance of mechanistic approaches highlights the central importance of physiological interpretability in cardiovascular digital twin development. Studies involving these approaches focus on replicating biophysical behavior with high fidelity, supporting diagnostic and interventional simulation tasks.

Data-driven twins, while less common, demonstrate growing interest in leveraging large clinical datasets for prediction, classification, and risk estimation. Their scope is more limited in scenarios requiring detailed physiological realism.

Hybrid methods illustrate emerging strategies that balance accuracy and computational efficiency. In studies involving these approaches, ML is commonly used to tune physiological parameters, derive boundary conditions from imaging or ECG data, or build surrogate models that reduce the computational cost of mechanistic solvers. In a subset of hybrid digital twin studies [51-56], ML components were typically embedded within a mechanistic framework rather than used in isolation. Across these studies, we observed 3 main integration patterns. First, ML is used for parameter tuning and personalization of mechanistic models, for example, by estimating subject-specific parameters or boundary conditions that are then supplied to a physics-based simulator. Second, ML algorithms are applied for feature extraction from raw clinical data, such as imaging or ECG signals, and the extracted features are subsequently used to initialize or constrain the mechanistic model. Third, in a small number of cases, ML serves as a surrogate or complementary model that approximates the behavior of a more complex mechanistic solver or is combined with mechanistic equations in a joint statistical-mechanistic framework. Together, these hybrid strategies illustrate how data-driven methods can enhance mechanistic digital twins by improving personalization, leveraging high-dimensional data, and reducing computational cost.

RQ2: If the Model is Mechanistic, What Specific Model Type is Used?

Overview

Across the 42 included studies, we identified multiple types of mechanistic models used within cardiovascular digital twin frameworks. Because individual studies often combined more than one formulation, we classified mechanistic components into 9 categories based on their predominant mathematical and physiological characteristics.

Key Insights

The key insights are as follows:

- Electrophysiology models were the most common (19 studies [11,17,18,20,21,25,27,28,30,34,35,37,39,41,43,44,46,51,54]). These models typically used monodomain, bidomain, or related reaction-diffusion formulations to simulate cardiac electrical activation, sometimes coupled to downstream mechanical effects.
- FEM-based structural or biomechanical models were used in 10 studies [8,18,19,26,27,29-31,33,37] to represent myocardial or vascular deformation, geometry, and stress-strain behavior.
- Electromechanical models, which explicitly couple electrical activation with tissue mechanics, were identified in 8 studies [22,26,30,32,33,52,53,55]. They supported integrated simulation of excitation-contraction processes.
- Simplified or system-level models, which are most often lumped-parameter formulations, were noted in 7 studies [24,29-31,33,36,53]. They provided compact descriptions of global hemodynamics or chamber-level dynamics, particularly when large-scale or long-duration simulations were required.
- Multiscale models were reported in 7 studies [11,22,30,33,41,44,53], linking processes across spatial or temporal scales (eg, from cellular electrophysiology to organ-level function).
- Computational fluid dynamics (CFD) models were used in 3 studies [40,42,55] to simulate blood flow and pressure distributions in chambers or great vessels. An additional 4 studies [19,23,24,56] used other mechanistic formulations (eg, specialized anatomical or biophysical models), and 1 study [42] used a surrogate mechanistic model that approximated a more complex solver. In 1 study [38], mechanistic modeling was reported, but the specific model type was not clearly described.

Interpretation

Taken together, the results show that electrophysiology-focused models form the backbone of mechanistic digital twin development in cardiology, with FEM-based structural models, lumped-parameter and multiscale formulations, and CFD models used in complementary roles. This diversity of model types illustrates how digital twin frameworks combine detailed biophysical fidelity with system-level abstractions to address specific clinical questions and RQs.

RQ3: If the Model Includes ML or AI, What Specific Algorithms are Applied?

Overview

Among the 42 reviewed studies, some explicitly reported using ML or AI techniques within the digital twin framework, while in others, the use or type of ML was absent or not clearly specified. Because several studies combined more than one method, we grouped algorithms into broad families, including DL, Bayesian methods, optimization algorithms, classical (statistical) ML, and other ML approaches.

Key Insights

The key insights are as follows:

- DL was the most frequently reported family of methods (9 studies [27,34,44-46,48,49,53,54]). These approaches included architectures, such as convolutional neural

XSL•FO
RenderX

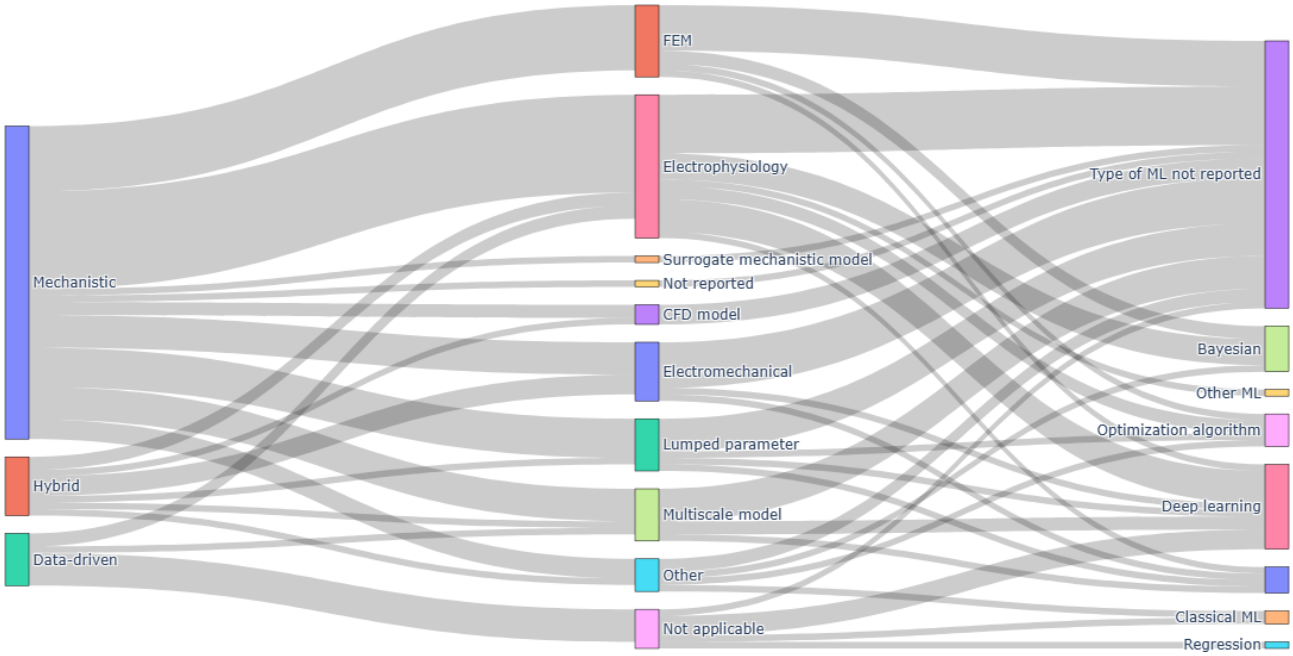
networks, neural operators, latent neural ordinary differential equation models, and related deep architectures, used for tasks like feature extraction, representation learning, or surrogate modeling.
- Bayesian methods were used in 5 studies [17,19,25,37,51], typically in the form of approximate Bayesian computation, Bayesian optimization, or Gaussian process–based models for probabilistic parameter estimation and uncertainty quantification.
- Optimization algorithms were noted in 4 studies [19,21,36,51]. These approaches included gradient-based schemes and metaheuristics that were used to tune model parameters, personalize simulations, or search over high-dimensional design spaces; in some cases, these optimizers were tightly integrated with Bayesian frameworks.
- Classical ML methods were identified in 2 studies [50,56]. These approaches included techniques, such as decision tree and logistic or linear regression, to model interpretable relationships between inputs and outcomes.
- Regression was explicitly highlighted as the primary approach in 1 of the studies [50]. One study used other ML strategies that did not fit neatly into the above categories but still relied on data-driven learning to support digital twin construction or personalization [46].

Interpretation

Overall, DL has emerged as the dominant explicitly reported ML family in cardiovascular digital twin research, supporting tasks such as feature extraction, surrogate modeling, and high-dimensional inference. Bayesian and optimization-based methods play a complementary role by enabling parameter estimation and uncertainty-aware personalization. Classical ML and regression, although less common, provide more interpretable models in selected use cases.

Figure 2 provides an integrated visualization of how primary modeling approaches, mechanistic model types, and ML or AI families co-occur across the included studies.

Figure 2. Relationships among modeling approaches, mechanistic model types, and machine learning (ML) or artificial intelligence (AI) methods in cardiovascular digital twin studies. Sankey diagram summarizing links among primary modeling approaches (research question [RQ] 1), mechanistic model types (RQ2), and ML or AI algorithm families (RQ3) across the 42 original research articles on cardiovascular digital twins included in this systematic review. The left column shows the dominant modeling approach for each study (mechanistic, hybrid, or data-driven). The middle column groups mechanistic model types into electrophysiology, finite element modeling (FEM), lumped parameter, electromechanical, computational fluid dynamics (CFD), other mechanistic models, and "not reported." The right column shows ML or AI families (deep learning, Bayesian methods, optimization algorithms, classical ML, regression, other ML, and "type of ML not reported"). The width of each flow is proportional to the number of studies combining the corresponding categories.



RQ4: Is the Framework or Model That is Created or Used Open-Source?

Overview

We evaluated the extent to which cardiovascular digital twin frameworks were shared as open-source resources. Code availability is a key indicator of scientific transparency and reproducibility, enabling independent validation and extension by other researchers and clinicians.

Key Insights

The key insights are as follows:

- Across the 42 included studies, 16 explicitly reported that their framework or model was available as open-source code [17-20,25,26,30,35,37,38,41,48,51,53,54,56].
- Two studies clearly stated that the code was not publicly released or that the implementation was proprietary [22,39].
- For the remaining 24 studies, code availability was either not mentioned or not described in sufficient detail to determine whether the implementation was accessible. Thus,

XSL•FO
RenderX

less than half of the studies (16/42, 38%) provided explicit evidence of open-source sharing, and in many cases, information on code availability was incomplete.

Interpretation

Despite increasing attention to reproducibility and open science, most studies in this review did not make their digital twin implementations publicly available. A lack of open-source code hinders transparency, reproducibility, and reusability. The few repositories that were shared provide valuable resources and serve as exemplars for future cardiovascular digital twin research.

Data Infrastructure and Visualization (RQ5 and RQ6)

The design and utility of cardiovascular digital twin systems depend heavily on how patient-specific data are structured, how outputs are visually communicated, and who the intended users are. This section addresses RQ5 and RQ6 by examining the types of data used to build or personalize digital twins, the formats used to present model outputs, and the target users of these visualizations. Together, these elements shaped the usability, interpretability, and clinical relevance of digital twin systems in practice.

RQ5: What Types of Patient-Specific Data are Used to Build or Personalize Digital Twins?

Overview

Patient-specific data underpin cardiovascular digital twin systems by enabling individual-level modeling. We explored the distinct categories of data used to personalize these models, ranging from electrical signals and anatomical imaging to omics and wearable-derived data. To facilitate interpretation, the data were grouped into consistent, semantically meaningful categories.

Key Insights

The key insights are as follows:

- Imaging data were the most commonly used (32 studies [8,11,17-19,21,22,24-27,29-31,33-40,42-44,46,48,49,51,53-55]). These data typically included modalities, such as MRI, CT, and other structural imaging, used to reconstruct patient-specific anatomy. Echocardiography was explicitly reported in 2 of these studies as a dedicated imaging source.
- Signal-based electrical data, primarily ECG, were used in 18 studies [8,17,21,25,27,33,36-39,43,46,48-51,53,54], reflecting its central role in modeling cardiac electrophysiology and conduction abnormalities.
- Vital signs were used in 12 studies [17,24,26,28,29,31,33,45,48,49,53,55], and demographics, such as age and sex, were reported in 9 studies [17,24,27,28,33,45-48], often to support model initialization, risk stratification, or cohort characterization.
- More detailed clinical information appeared in several categories: omics data were used in 4 studies [22,23,49,55], lab results were used in 4 studies [22,28,49,55], and general clinical data (such as clinical histories and visit summaries) were used in 3 studies [49,52,55]. Diagnosis [33,47,48] and treatment-related data [33,47,48] (eg, information on interventions or therapies) were each reported in 3 studies.
- Sensor-based and longitudinal monitoring information was less common: 3 studies used data from sensors [46,49,55], and 2 studies used activity tracker data [45,49], illustrating the early integration of wearable or home-based measurements into digital twin personalization. Synthetic patient data were explicitly used in 1 study [56].

Interpretation

Overall, there is a strong reliance on imaging and ECG data to define anatomy and electrophysiological behavior in cardiovascular digital twins, complemented by vital signs and demographic information for basic personalization. Omics, lab results, richer clinical records, and wearable or sensor-derived data are beginning to appear but remain less common, suggesting that truly multimodal, longitudinal personalization is still emerging. The presence of synthetic and other less conventional data sources indicates ongoing experimentation with alternative data strategies.

RQ6: What is the Primary Format Used to Visually Present Digital Twin Outputs?

Overview

We examined how digital twin outputs were visualized in cardiovascular studies, an essential aspect for interpretation, user interaction, and eventual clinical integration. Each study could use more than one visualization format, so outputs were classified into standard categories such as static figures, anatomical renderings, tables, dashboards, and interactive media.

Key Insights

The key insights are as follows:

- Static figures were the most common visualization format (41 studies [8,11,17-47,49-56]). These typically included plots, error curves, comparative graphics, and screenshots of simulations, and were primarily designed for inclusion in scientific publications.
- Two- or three-dimensional anatomical views were reported in 27 studies [8,11,17-19,21,22,24-27,29-31,33-35,37,39,40,42-44,46,51,54,55], where patient-specific geometries or simulated fields (eg, activation times, strain, and flow patterns) were mapped onto cardiac or vascular structures. These views served to visually link model predictions to underlying anatomy.
- Tabular formats were used in 7 studies [19,22,24,35,36,45,49] to report numerical outputs such as performance metrics, parameter values, and summary statistics.

Interpretation

Overall, visualization of cardiovascular digital twins remains dominated by static, publication-oriented formats such as figures and anatomical snapshots, with limited support for dynamic, interactive, or dashboard-based exploration. While anatomical views help contextualize outputs in patient-specific geometry, the scarcity of dashboards, animations, and interactive interfaces suggests that user-centric and real-time visualization capabilities are still underdeveloped. Enhancing interactive and clinically oriented visualization tools may be crucial for translating digital

XSL•FO
RenderX

twins from research prototypes into practical decision-support systems.

Clinical Applications (RQ7 and RQ8)

We explored how digital twins were applied in clinical cardiology (RQ7) and which cardiovascular conditions they targeted (RQ8). It highlighted current use cases, such as diagnosis, planning, and monitoring, and categorized the conditions based on thematic grouping identified during full-text analysis.

RQ7: What is the Main Clinical Application or Use Case of Digital Twin Systems?

Overview

We explored the primary clinical applications of cardiovascular digital twin systems, revealing the core motivations behind their development and deployment. Use cases ranged from therapy planning and risk prediction to monitoring, drug testing, and more exploratory clinical applications. Individual studies could contribute to multiple application categories.

Key Insights

The key insights are as follows:

- Therapy planning was the most common application (28 studies [8,11,17,19,20,22-25,29-34,36-39,41,45-47,49,51,54-56]). In these studies, digital twins were used to support the selection, personalization, or optimization of interventions, including device configuration, ablation strategies, or other patient-specific treatment plans.
- Risk prediction was noted in 11 studies [20,28,40,41,46-50,52,55], where digital twins were used to estimate the likelihood of adverse events, treatment responses, or disease trajectories, often to support patient stratification. Diagnosis-focused applications were identified in 7 studies [11,45,46,48-50,54], using digital twins to assist in identifying underlying pathophysiology or classifying clinical conditions.
- Surgical and device simulation was reported in 6 studies [36,38,42,46,51,55], in which digital twins provided virtual testbeds to explore procedural strategies or evaluate device performance in patient-specific anatomies. Another 6 studies used digital twins for drug testing [17,20,28,32,36,37].
- Monitoring applications were noted in 6 studies [45,48-50,52,55], where digital twins contributed to disease tracking or follow-up by integrating longitudinal data or repeated assessments. Disease progression modeling was explicitly highlighted in 3 studies [36,42,55], and a single study focused primarily on prognosis [55].

Interpretation

Overall, cardiovascular digital twins are most frequently positioned as tools for therapy planning and risk prediction, emphasizing their role in personalizing and optimizing clinical interventions. Diagnosis, surgical or device simulation, drug testing, and monitoring collectively demonstrate a broad range of applications along the care pathway, from early risk assessment to procedural planning and follow-up. As digital twin technologies mature, a clearer definition and reporting of clinical applications will be important for understanding their real-world impact.

RQ8: What Cardiovascular Conditions Are Studied Using Digital Twin Systems?

Overview

We examined the range of cardiovascular conditions addressed by digital twin systems, providing a disease-centered perspective on where digital twin technologies are currently being applied. Conditions were grouped into clinically meaningful categories, and the classification was reviewed by a physician on the research team to ensure clinical relevance and consistency.

Key Insights

The key insights are as follows:

- Arrhythmia was the most frequently studied condition (13 studies [8,11,18,20,25,28,30,34,38-41,43]). The studies predominantly focused on atrial fibrillation and other rhythm disorders, reflecting the suitability of digital twins for simulating electrophysiological mechanisms and guiding rhythm-related interventions.
- Heart failure was investigated in 9 studies [33,36,38,41,48,51-53,55], often in the context of global cardiac function, ventricular remodeling, or device-based therapies. Cardiomyopathies, including hypertrophic cardiomyopathy and other structural myocardial diseases, were the primary focus in 5 studies [19,22,32,35,44], where digital twins were used to explore patient-specific mechanics and electrophysiology.
- Six studies centered on healthy or control populations [17,43,46,49,54,56], using digital twins to represent normal physiology, establish reference behaviors, or provide baselines for comparison with diseased states. Aortic disease was the focus in 3 studies [26,29,42], typically involving patient-specific modeling of the aorta for flow, wall stress, or device evaluation.

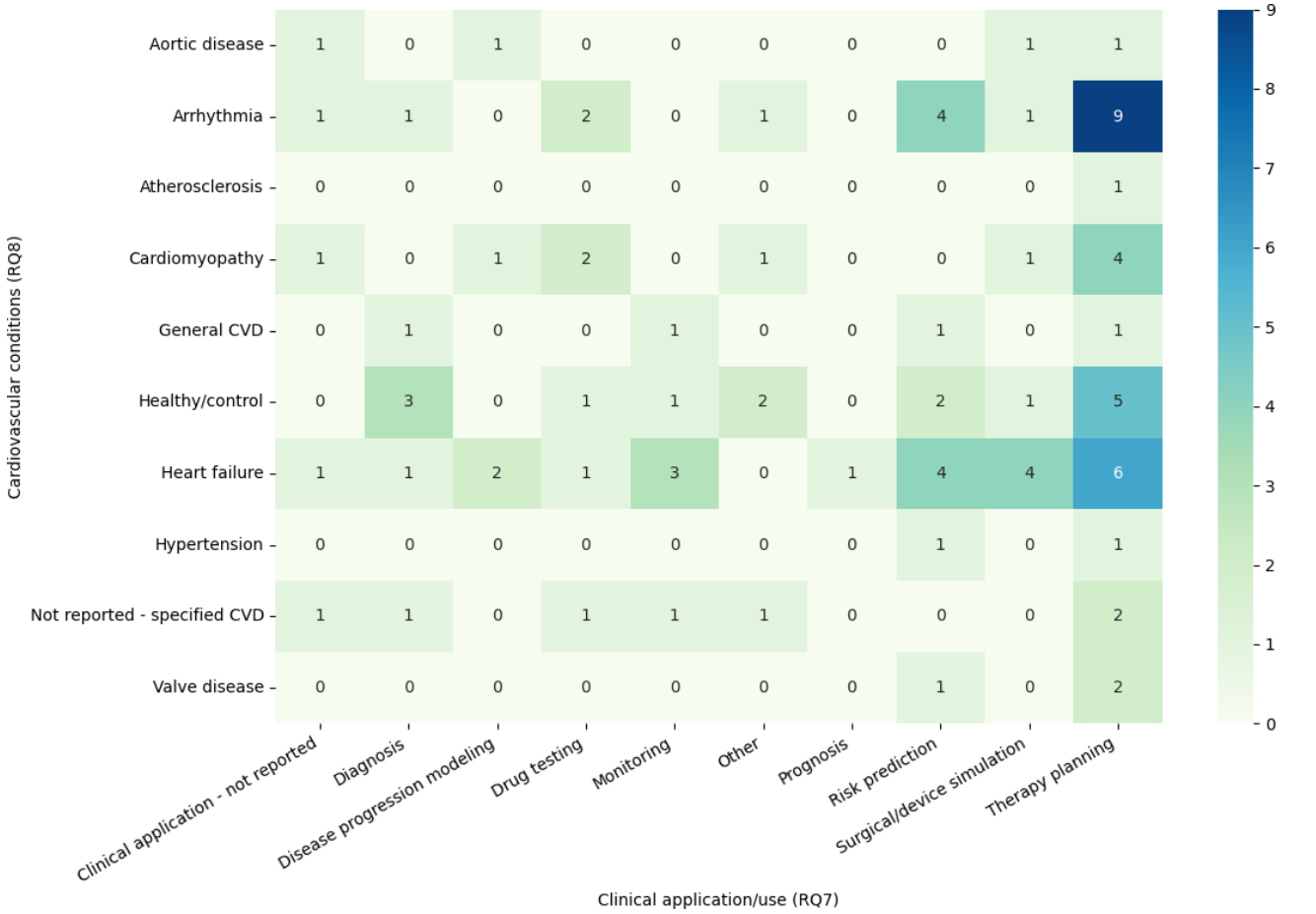
Interpretation

The distribution shows a strong emphasis on arrhythmia and heart failure, conditions in which digital twins can leverage detailed electrophysiological and hemodynamic modeling to support diagnosis, therapy planning, and risk assessment. Cardiomyopathies, aortic disease, and valvular disease are also emerging areas of application, particularly where structural and flow abnormalities can be represented in patient-specific models. By contrast, hypertension, atherosclerosis, and some other common cardiovascular conditions are only sporadically represented, and several studies do not clearly specify the underlying disease focus. These gaps suggest opportunities for expanding digital twin applications into a broader spectrum of cardiovascular conditions and for improving the clarity of disease reporting in future work.

Figure 3 summarizes how clinical applications are distributed across cardiovascular conditions. As shown in Figure 3, therapy planning and risk prediction are concentrated in arrhythmia and heart failure, whereas other conditions and applications are represented by only a small number of studies, underscoring the uneven distribution of digital twin work across CVDs.

XSL•FO
RenderX

Figure 3. Heatmap of cardiovascular conditions (research question [RQ] 8) versus clinical applications (RQ7) in cardiovascular digital twin studies. Rows show the primary cardiovascular condition modeled (eg, arrhythmia, heart failure, cardiomyopathy, aortic and valve disease, hypertension, atherosclerosis, general cardiovascular disease [CVD], healthy/control, and not reported). Columns show the main clinical applications (eg, diagnosis, disease progression modeling, drug testing, monitoring, prognosis, risk prediction, surgical or device simulation, and therapy planning). Cell color and numbers indicate how many of the 42 included studies reported each condition-application combination (darker cells indicate a higher number of studies).



Impact on Clinical Practice (RQ9)

We identified the reported clinical benefits of cardiovascular digital twin systems, including improved accuracy, personalization, therapy planning, and patient outcomes. The findings were organized into key impact categories to highlight where digital twins showed practical value in care delivery.

RQ9: What Clinical Impacts are Reported as a Result of Using Digital Twins?

Overview

We examined the concrete clinical or clinically relevant impacts attributed to cardiovascular digital twin systems. Rather than focusing on intended use alone, we captured reported effects where the use of a digital twin was described as influencing decision-making, therapy, diagnostic performance, or other aspects of care. Reported impacts were grouped into categories such as improved decision-making, therapy-related benefits, increased accuracy, and other specific outcomes.

Key Insights

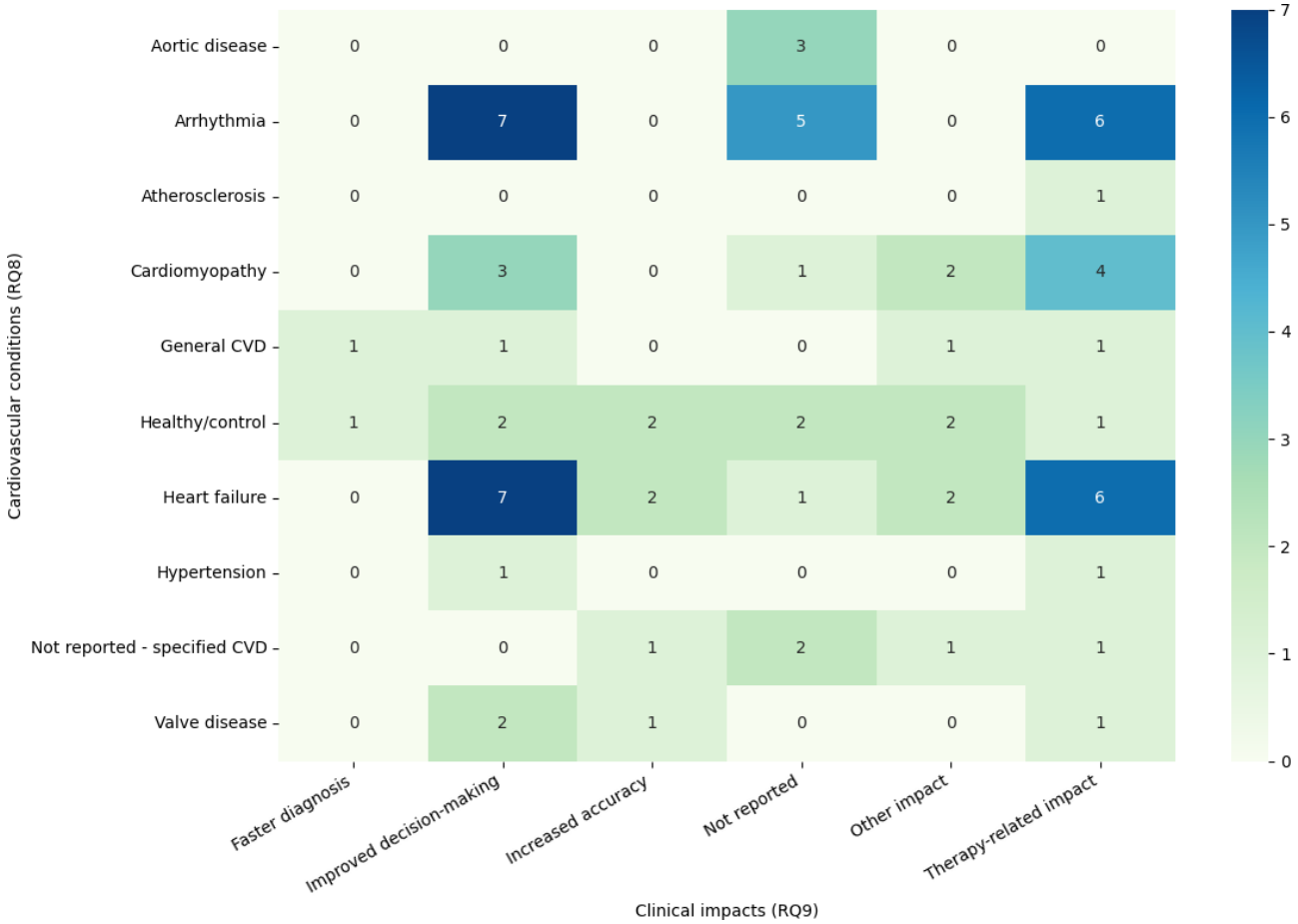
The key insights are as follows:

- Improved decision-making was the most frequently reported impact (19 studies [22,24,28,30-34,36,38-41,47-49,51,55,56]). In these cases, digital twins were described as helping clinicians compare alternative strategies, understand patient-specific mechanisms, or select interventions with greater confidence.
- Therapy-related impacts were reported in 18 studies [19,22-25,32,34,36,38-41,45,47-49,51,55], including optimization of device settings, refinement of ablation targets, adjustment of pharmacologic regimens, and more tailored procedural planning based on virtual simulations.
- Increased accuracy was explicitly identified in 6 studies [31,45,48,49,54,55], referring to improvements in predictive performance, better correspondence between simulations and measured clinical data, or more faithful reproduction of patient-specific physiology. Two studies [49,50] reported a faster diagnostic process, where digital twin–supported workflows were associated with quicker identification or clarification of clinical conditions.

Figure 4 illustrates how reported clinical impacts are distributed across cardiovascular conditions. Improved decision-making and therapy-related impacts were concentrated in arrhythmia and heart failure, whereas many other condition-impact combinations were represented by only one or two studies, highlighting the uneven evidence base across disease areas.

XSL•FO
RenderX

Figure 4. Heatmap of cardiovascular conditions (research question [RQ] 8) versus reported clinical impacts (RQ9) in cardiovascular digital twin studies. Rows represent the primary cardiovascular condition modeled by the digital twin (eg, arrhythmia, heart failure, cardiomyopathy, aortic and valve disease, hypertension, atherosclerosis, general cardiovascular disease [CVD], healthy/control, and not reported). Columns represent impact categories reported by study authors (faster diagnosis, improved decision-making, increased accuracy, impact not reported, other impact, and therapy-related impact). Cell color and numbers indicate how many of the 42 included studies reported each condition-impact combination (darker cells indicate a higher number of studies).



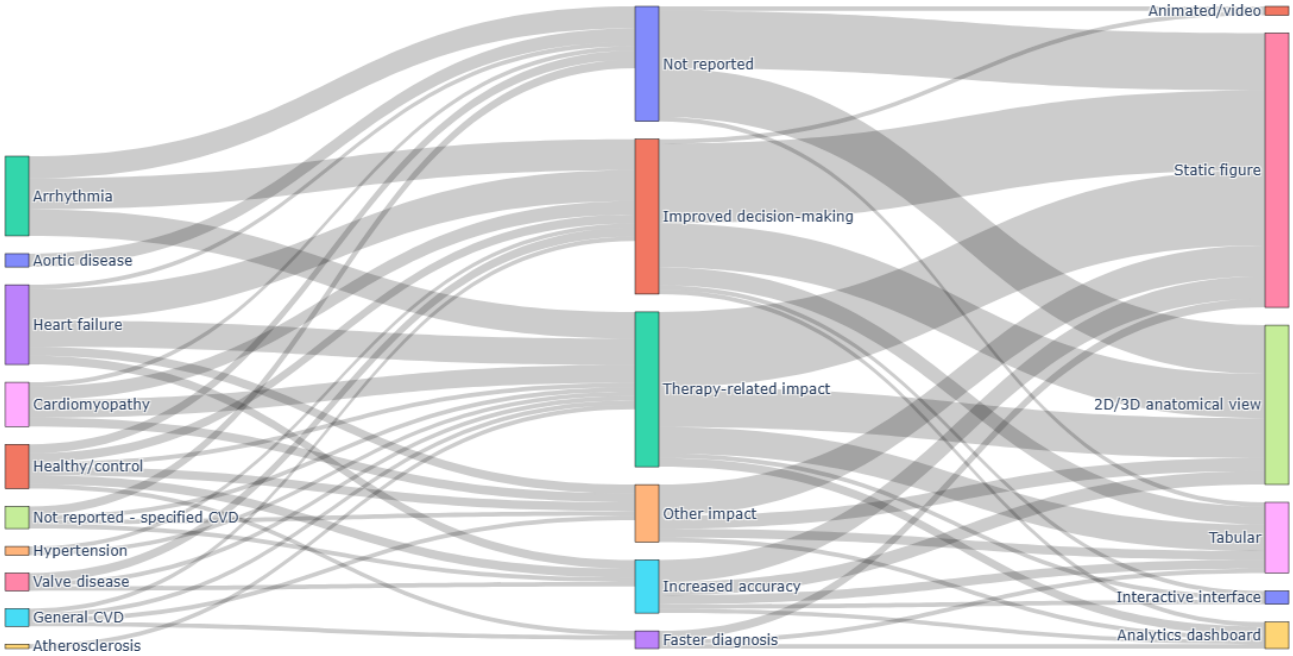
Interpretation

The most commonly reported benefits of cardiovascular digital twins relate to improved clinical decision-making and therapy-related impacts, suggesting that these systems are beginning to influence how clinicians choose and personalize interventions. Explicit gains in accuracy and diagnostic speed are less frequently reported but point toward the quantitative advantages of model-based approaches when they are carefully evaluated. At the same time, the substantial number of studies with no clearly articulated clinical impact indicates that much of the current literature remains focused on technical feasibility and validation rather than demonstrated downstream effects on care processes or patient outcomes. Strengthening the evidence base around measurable clinical benefits, such as improved decision quality, optimized therapy, and better outcomes, will be essential for wider clinical adoption.

Figure 5 shows that improved decision-making is the dominant reported impact across most cardiovascular conditions, particularly heart failure and arrhythmia, and that these impacts are almost always communicated through static figures and 2D or 3D anatomical views rather than dashboards, animations, or interactive interfaces. Therapy-related impacts and gains in accuracy are more sparsely reported and similarly rely on conventional publication-style visualizations, underscoring the limited development of user-facing, real-time visual tools, even in high-risk clinical scenarios.

XSL•FO
RenderX

Figure 5. Relationships among cardiovascular conditions, reported clinical impacts, and visualization formats in cardiovascular digital twin studies. Sankey diagram summarizing links among cardiovascular conditions (research question [RQ] 8), reported clinical impacts (RQ9), and primary visualization formats (RQ6) across the 42 original research articles on cardiovascular digital twins included in this systematic review. The left column shows the main conditions modeled by the digital twins (eg, heart failure, arrhythmia, valve disease, cardiomyopathy, hypertension, atherosclerosis, general cardiovascular disease [CVD], and healthy/control populations). The middle column displays impact categories reported by the authors (eg, improved decision-making, therapy-related impact, increased accuracy, faster diagnosis, and other impact). The right column shows the dominant visualization formats used to present model outputs (static figures, 2D/3D anatomical views, tabular displays, analytics dashboards, animated/video outputs, and interactive interfaces). The width of each flow is proportional to the number of studies combining the corresponding categories.



Barriers to Implementation and Ethical Considerations (RQ10 and RQ11)

We examined the key challenges limiting the adoption of cardiovascular digital twins, including technical barriers (RQ10) and ethical or legal concerns (RQ11). These issues highlighted the need for improved scalability, transparency, and responsible use in clinical settings.

RQ10: What Limitations or Practical and Technical Barriers are Described?

Overview

We identified the limitations and implementation barriers of cardiovascular digital twin systems as reported by the included studies. Rather than listing every individual issue, reported limitations were grouped into conceptually meaningful categories, such as model assumptions, computational constraints, data-related challenges, and integration or usability problems. This categorization helped highlight systemic obstacles that recur across the field.

Key Insights

The key insights are as follows:

- Model assumptions and structural simplifications were the most frequently cited limitations (26 studies [8,11,17,19-22,24,25,27-33,35,36,39,41,43,44,46,47,51,53]). These concerns included oversimplified anatomy or physiology, restrictive boundary conditions, and reduced model complexity that may limit generalizability or omit important mechanisms.
- Computational cost was highlighted in 21 studies [11,17,21,22,26,27,29-31,33,36,37,39,41,42,46,49,51,53,54,56], where authors noted long simulation times, high hardware requirements, or overall computational burden that can impede large-scale studies and real-time or near–real-time clinical use.
- Data-related challenges were prominent, with 16 studies [19,22,25,27-30,36,39,40,43,45,47,50,55,56] reporting issues with data quality or availability, such as incomplete or noisy clinical inputs, limited access to high-resolution or longitudinal data, and difficulties in acquiring truly personalized datasets. Limited validation was also mentioned in 16 studies [24,25,27-29,31,33-36,39,40,45,47,53,55], reflecting concerns about small sample sizes, restricted cohorts, synthetic data, or a lack of robust testing in real-world clinical environments.
- More specific barriers included a lack of real-time performance in 5 studies [33,42,50,54,55], indicating that even when models were accurate, their latency or compute demands were not compatible with time-sensitive clinical workflows. Workflow integration problems were identified in 4 studies [30,47,51,55], focusing on the challenges of embedding digital twins into existing clinical systems and processes. Clinician usability challenges were noted in 3 studies [45,49,55], where interfaces or outputs were considered difficult to interpret or not well aligned with clinical practice. High infrastructure cost was noted in 2 studies [49,53], and data security or privacy concerns were explicitly mentioned in 1 study [49].

XSL•FO
RenderX

Interpretation

The most common limitations—strong model assumptions, high computational cost, and data and validation constraints—reflect the technical and methodological complexity of deploying cardiovascular digital twins in practice. Simplifying assumptions and limited data can undermine generalizability, while computational burden and lack of real-time performance can restrict clinical usability. Integration issues, usability challenges, infrastructure demands, and security concerns, though mentioned less often, highlight important practical barriers that will become more pressing as digital twins move closer to clinical deployment. Addressing these limitations through improved model design, better data infrastructure, efficient algorithms, and user-centered integration will be essential for scalable, clinically viable digital twin systems.

RQ11: What Legal, Ethical, or Data Governance Issues are Raised Regarding Digital Twins?

Overview

We explored the ethical, legal, and data governance concerns raised in studies involving cardiovascular digital twin systems. Potential issues included privacy protection, regulatory compliance, informed consent, algorithmic transparency, and fairness. Reported concerns were grouped into categories to highlight common themes and gaps in current practice.

Key Insights

The key insights are as follows:

- Only a small subset of studies explicitly discussed legal, ethical, or governance issues. Privacy and data protection were the most frequently mentioned topics, identified in 4 studies [47,49,50,55], with references to compliance frameworks, such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), and concerns about safeguarding sensitive patient data in the context of high-dimensional digital representations.
- Some studies raised other specific issues. Two studies discussed ethical or legal challenges in general terms [30,47], while another identified potential algorithmic bias, described problems or open questions around informed consent, and highlighted concerns about model transparency and the need for explainable or interpretable digital twin behavior [55].

Interpretation

Overall, explicit discussion of legal, ethical, and data governance aspects remains limited in the cardiovascular digital twin literature. While privacy and regulatory compliance are beginning to appear as concrete concerns, far fewer studies engage with broader questions around algorithmic bias, transparency, informed consent in the context of complex modeling, or downstream legal responsibilities. As digital twin systems move closer to clinical deployment and real-world decision support, more systematic attention to these dimensions, including fairness, accountability, liability, and data stewardship, will be critical to ensure trustworthy and responsible adoption.

Risk of Bias Assessment

The risk of bias was assessed for all included studies using the tool corresponding to the underlying study design. Among the 42 studies evaluated, 38 were computational or simulation-based studies assessed using the custom modeling checklist [8,11,17,18,19,20,21-27,28,29,30,31,32-34,35,36,37,38,40,41,42,43,44,45,46,48,51-53,54,55,56], 2 were prediction-modeling studies evaluated using the PROBAST [49,50], and 2 were observational cohort studies evaluated using the ROBINS-I [39,47].

Table 2 summarizes the distribution of overall risk-of-bias judgments across the 3 tools. For simulation and digital twin modeling studies, “unclear” was the most frequent overall rating (22/38, 58%), followed by “high risk” (16/38, 42%). The domains contributing most frequently to elevated risk included data representativeness, validation strategy, and sample size/overfitting. No modeling study received an overall low-risk judgment, reflecting commonly observed methodological limitations in data availability, external validation, and reproducibility practices across computational literature.

Table . Summary of overall risk-of-bias judgments across the included studies.

Tool	Total studies (N=42), n	Unclear risk, n (%)	High risk, n (%)
Custom modeling checklist	38	22 (58)	16 (42)
PROBAST[a]	2	0 (0)	2 (100)
ROBINS-I[b]	2	1 (50)	1 (50)

[a]PROBAST: Prediction Model Risk of Bias Assessment Tool.

[b]ROBINS-I: Risk of Bias in Non-Randomized Studies - of Interventions.

Both prediction-modeling studies assessed with the PROBAST were rated as having a high risk of bias, predominantly due to concerns in the analysis and outcome domains, including insufficient handling of model calibration, unclear predictor specification, and absence of prespecified analysis protocols.

Among the 2 observational cohort studies evaluated using the ROBINS-I, one was judged as having a high risk of bias, primarily due to serious confounding and selective reporting, while the other was rated as unclear.

Figures 6 and 7 present the traffic-light and summary plots, respectively, for all 38 modeling studies assessed using the custom modeling checklist. These visualizations highlight consistent methodological limitations across key domains, particularly external validation and representativeness of data inputs. Traffic-light and summary plots for the PROBAST and

XSL•FO
RenderX

ROBINS-I assessments are provided in Figures 8 and 9, respectively.

A structured visualization workflow was implemented using the robvis tool, which standardizes the graphical representation of domain-level and overall judgments and supports transparent reporting of risk-of-bias evaluations.

XSL•FO
RenderX

Figure 6. Risk of bias assessment (traffic-light plot) for modeling studies [8,11,17,18,19,20,21-27,28,29,30,31,32-34,35,36,37,38,40,41,42,43,44,45,46,48,51-53,54,55,56]. Traffic-light plot for the 38 simulation/digital twin modeling studies assessed using the custom modeling checklist. Domain-level judgments are categorized as low, unclear, or high. The plot has been generated using the robvis tool [16].

XSL•FO
RenderX

Figure 7. Risk of bias assessment (summary plot) for modeling studies [8,11,17,18,19,20,21-27,28,29,30,31,32-34,35,36,37,38,40,41,42,43,44,45,46,48,51-53,54,55,56]. Summary plot for the 38 simulation/digital twin modeling studies assessed using the custom modeling checklist. Domain-level judgments are categorized as low, unclear, or high. The plot has been generated using the robvis tool [16].

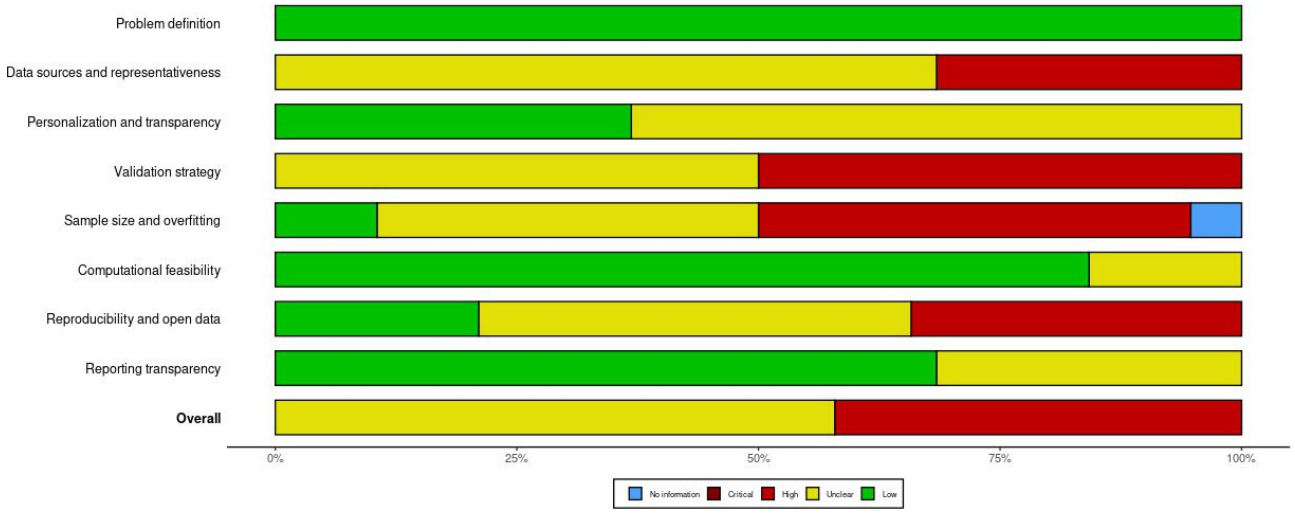
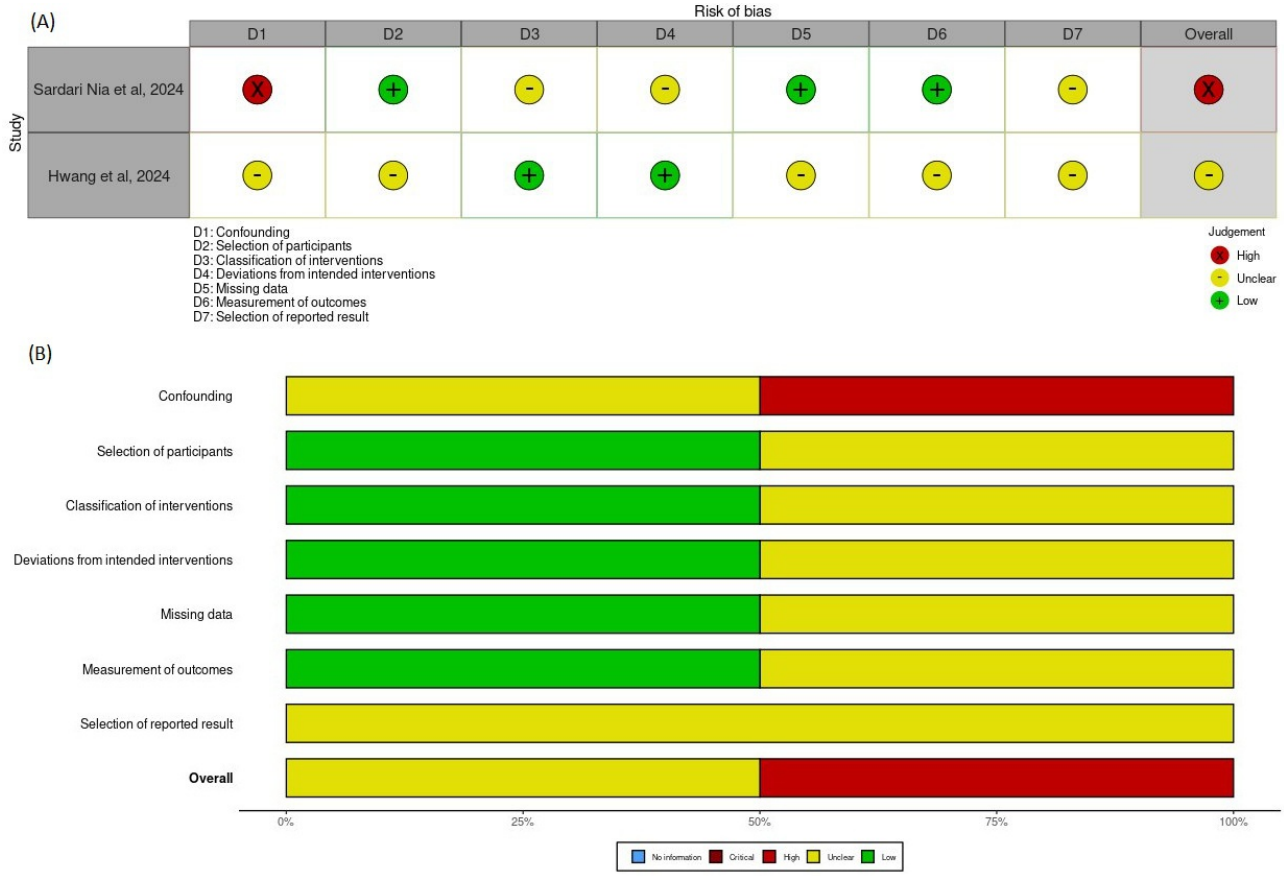


Figure 8. Risk of bias assessment for prediction-modeling studies (Prediction Model Risk of Bias Assessment Tool [PROBAST]) [49,50]. Traffic-light plot (A) and summary plot (B) for the 2 prediction-modeling studies evaluated using the PROBAST instrument. Judgments are shown across the 4 PROBAST domains (participants, predictors, outcome, and analysis) and the overall study-level rating. Visualizations are created using the robvis tool [16].



XSL•FO
RenderX

Figure 9. Risk of bias assessment for observational cohort studies (Risk of Bias in Non-Randomized Studies - of Interventions [ROBINS-I]) [39,47]. Traffic-light plot (A) and summary plot (B) for the 2 observational cohort studies evaluated using the ROBINS-I tool. Judgments are shown across the 7 ROBINS-I bias domains and the overall risk-of-bias rating. Visualizations are created using the robvis tool [16].



Discussion

Principal Findings

This systematic review synthesized findings from 42 studies and showed that cardiovascular digital twin technology is progressing rapidly but remains largely preclinical and methodologically heterogeneous. Most systems relied on mechanistic models, with a smaller subset incorporating explicit ML or hybrid mechanistic–data-driven designs. Applications clustered around arrhythmia (13/42, 31%), heart failure (9/42, 21%), and therapy planning (28/42, 67%), yet relatively few studies reported real-world clinical deployment, rigorous validation (16/42, 38%), or patient-level outcomes, underscoring the gap between technical innovation and routine clinical use.

Across 11 RQs spanning modeling foundations, data infrastructure, clinical applications, clinical impact, and implementation challenges, the review identified steady technical progress alongside persistent limitations in data quality, external validation, usability, and ethical governance. Our structured risk-of-bias assessment further highlighted that most modeling and prediction studies were judged as having unclear or high risk of bias, particularly in relation to data representativeness, validation strategies, and analysis procedures. Together, these findings suggest that cardiovascular digital twins are scientifically promising but not yet ready for widespread clinical translation.

Technological Foundations and Modeling Strategies

Mechanistic models form the backbone of current cardiovascular digital twins. Electrophysiology, finite-element structural modeling, lumped-parameter formulations, multiscale frameworks, and CFD-based flow simulations were frequently combined to capture different physiological scales and processes. The predominance of mechanistic approaches reflects the central importance of physiological interpretability and explicit biophysical assumptions in cardiology, where understanding causal mechanisms is often as important as prediction performance.

Hybrid designs and explicit ML or AI integrations were present but less common than might be expected given the broader trends in digital health. Only a minority of studies (18/42, 43%) clearly described ML algorithms, with DL (9/42, 21%), Bayesian methods (5/42, 12%), and optimization algorithms (4/42, 10%) used for tasks such as parameter estimation, feature extraction, surrogate modeling, and uncertainty quantification. Many other papers referred to “ML” or “AI” without specifying algorithm families or training procedures, limiting reproducibility and comparability. Open-source dissemination was also limited; less than half of the studies (16/42, 38%) provided accessible code, constraining independent verification, reuse, and benchmarking.

Data Infrastructure and Visualization

Personalization of cardiovascular digital twins relied heavily on structural imaging (32/42, 76%) and electrical signals (18/42,

XSL•FO
RenderX

43%). Imaging data (most often MRI or CT, with occasional echocardiography) enabled patient-specific anatomical reconstruction, while ECG and related electrical measurements supported modeling of activation patterns and conduction abnormalities. Vital signs (12/42, 29%) and demographic variables (9/42, 21%) were commonly used as basic covariates, but richer data sources appeared in only a subset of studies (omics: 4/42, 10%; lab results: 4/42, 10%; detailed clinical records: 3/42, 7%; and wearable/sensor streams: 4/42, 10%). This pattern suggests that many digital twins remain anchored in traditional imaging and electrophysiology pipelines, with multimodal, longitudinal data integration still in an early stage.

Visualization practices were predominantly static and publication-oriented. Most studies communicated digital twin outputs through static figures (41/42, 98%), anatomical overlays (27/42, 64%), or tables summarizing simulation results (7/42, 17%). Only a few described dashboards, dynamic animations, or interactive interfaces that would support real-time exploration or clinical decision-making. As a result, the "front end" of many digital twin systems remains geared toward researchers rather than clinicians or patients, which may hinder adoption even when the underlying models are sophisticated.

Clinical Applications and Target Conditions

Clinically, cardiovascular digital twins were most frequently positioned as tools for therapy planning (28/42, 67%), risk prediction (11/42, 26%), and monitoring (6/42, 14%), with additional roles in diagnosis (7/42, 17%), surgical or device simulation (6/42, 14%), and drug testing (6/42, 14%). Conditions, such as atrial fibrillation and other arrhythmias (13/42, 31%), heart failure (9/42, 21%), cardiomyopathy (5/42, 12%), and aortic disease (3/42, 7%), were most commonly represented, reflecting both their high burden and the suitability of these conditions for simulation-based assessment. Several studies (6/42, 14%) used digital twins to model healthy or control populations, providing physiological baselines and enabling comparison with diseased states.

At the same time, important cardiovascular domains remain underrepresented. Hypertension, atherosclerosis, congenital heart disease, and some valvular pathologies appeared relatively rarely or were only indirectly addressed, despite their major contribution to global cardiovascular morbidity. Furthermore, in some studies (4/42, 10%), the underlying clinical condition was not clearly specified, blurring the line between generic modeling exercises and disease-focused digital twin applications. This uneven coverage limits our ability to generalize digital twin findings across the broader spectrum of CVD.

Impact on Clinical Practice

Reported clinical impacts aligned with the conceptual promise of digital twins but were often indirect or inferred. The most commonly cited benefits were improved decision-making (19/42, 45%) and therapy-related impacts (18/42, 43%), including better selection of interventions, refined device configurations, and more personalized procedural planning. Some studies (6/42, 14%) reported increased accuracy of predictions or simulations, and a small number of studies (2/42, 5%) documented faster diagnosis or workflow advantages.

However, very few studies (4/42, 10%) linked digital twin use to robust patient-level outcomes such as mortality, hospitalization, and long-term symptom burden. Most evidence came from retrospective analyses, in silico comparisons, or small proof-of-concept applications rather than prospective, real-world evaluations. Consequently, while digital twins appear to enhance mechanistic understanding and may plausibly improve decision quality, the causal pathway from digital twin use to improved patient outcomes remains largely hypothetical. This observation was reinforced by the risk-of-bias assessments, which highlighted frequent limitations in sample size, external validation, and outcome measurement.

Barriers to Implementation and Ethical Considerations

Several recurring barriers emerged across the included studies. Strong model assumptions and structural simplifications, while often necessary for tractability, raise questions about generalizability to broader populations or clinical settings. High computational cost and limited real-time performance constrain scalability and integration into time-sensitive workflows, particularly in acute care or interventional environments. Data-quality issues, including incomplete or noisy inputs and limited access to comprehensive, longitudinal datasets, further restrict personalization and increase uncertainty.

Workflow integration and clinician usability remain significant challenges. Only a minority of studies (4/42, 10%) described how digital twin systems might be embedded within electronic health records, imaging systems, or existing decision-support tools, and even fewer studies (3/42, 7%) reported formal usability testing with clinicians. Ethical, legal, and governance issues were discussed explicitly in only a small subset of articles (4/42, 10%), primarily in relation to privacy and data protection frameworks such as GDPR and HIPAA. Isolated studies mentioned algorithmic bias, informed consent, or transparency concerns (1/42, 2%), but systematic engagement with liability, accountability, data ownership, and equity was rare, despite their importance for future clinical deployment.

Sources and Implications of Heterogeneity

Across the included studies, we observed substantial heterogeneity in how cardiovascular digital twins were conceptualized, implemented, and evaluated. This variability spanned multiple dimensions, including the definition and scope of the "digital twin," the underlying modeling strategies (eg, electrophysiology, finite-element, lumped-parameter, multiscale, CFD, and hybrid ML-mechanistic designs), the types and combinations of data modalities used for personalization, the clinical applications and disease targets, and the choice of validation approaches and outcome metrics. As a result, the findings are difficult to compare directly across studies, and a quantitative synthesis or meta-analysis is not appropriate. This heterogeneity also limits the generalizability of individual results and makes it challenging to derive standardized performance expectations for cardiovascular digital twins. Future work will benefit from clearer definitions, minimum reporting standards, and shared benchmarks to enable a more systematic comparison and aggregation of evidence.

XSL•FO
RenderX

Implications and Future Directions

The findings of this review suggest that cardiovascular digital twins are technically promising but not yet consistently validated, standardized, or integrated into routine care. Heterogeneity in modeling approaches, data inputs, validation strategies, and reporting practices limits comparability and makes it difficult to draw firm conclusions about real-world effectiveness.

Future work should focus on strengthening clinical validation in real-world settings, ideally through prospective and multisite studies that link digital twin use to patient outcomes and workflow changes. In parallel, clearer definitions of what constitutes a digital twin; shared performance metrics; and minimum reporting standards for models, data, and validation would support meaningful comparisons and regulatory assessments. Methodological transparency and user-centered design are also essential. Explainable or interpretable modeling pipelines and clinician-oriented interfaces are likely to be critical for trust and adoption.

Finally, ethical and equity considerations need to be addressed proactively. Most existing studies draw on narrow populations and rarely examine algorithmic bias, informed consent for complex modeling, or long-term data governance. Future research should deliberately include diverse populations and care settings; evaluate generalizability across subgroups; and embed privacy protection, transparency, and fair data use into the design and deployment of digital twin systems. Closer collaboration with regulators and health care organizations will be important to ensure that these technical and ethical advances translate into safe, accountable, and clinically useful tools.

Limitations of This Review

While comprehensive, this review may have missed relevant studies, especially studies published in non-English sources or proprietary implementations outside academic literature. Reporting heterogeneity also limited the comparability of validation and outcome data. As the field evolves rapidly, some emerging developments may not have been captured in the included studies.

Although we conducted a structured risk-of-bias appraisal using a custom modeling checklist for simulation studies, the PROBAST for prediction models, and the ROBINS-I for observational cohort studies, these tools were not originally designed for all types of cardiovascular digital twin research and required judgment-based adaptation. In addition, the substantial heterogeneity in study designs, data sources, and evaluation strategies precludes quantitative synthesis and indicates that our risk-of-bias judgments should be interpreted as broad indicators of methodological robustness rather than definitive ratings for individual studies.

This review was based on searches of major bibliographic databases and Google Scholar but did not include dedicated, systematic searches of clinical trial registries (eg, ClinicalTrials.gov), conference proceedings, or specialized grey-literature repositories (eg, dissertation or technical report databases). Although Google Scholar can index some gray literature and conference outputs, our screening was not designed to comprehensively capture these sources. As a result, ongoing trials, conference-only presentations, and nontraditionally published or proprietary digital twin implementations may be underrepresented in this synthesis.

Finally, the review protocol was not registered on a public platform, such as the Open Science Framework (OSF), which may limit reproducibility and transparency. Future work would benefit from prospective protocol registration to reduce the risk of selective reporting and enhance methodological rigor.

Conclusion

This systematic review mapped the technological, clinical, and implementation landscape of cardiovascular digital twin systems across 42 original studies. We found that most digital twins are grounded in mechanistic modeling, with limited but growing use of hybrid and AI-driven approaches. Personalization relies predominantly on imaging and electrical signals, and applications are concentrated in therapy planning, risk prediction, and monitoring for arrhythmia and heart failure. Although the reported impacts on decision-making and therapy optimization are promising, evidence for downstream patient-level benefits remains sparse.

At the same time, substantial heterogeneity in model architectures, data modalities, clinical use cases, and validation strategies—combined with incomplete reporting of algorithms, data, and code—limits comparability across studies and precludes quantitative synthesis. Key barriers to clinical translation include strong modeling assumptions; high computational cost; constrained data quality and availability; and limited real-time performance, workflow integration, and usability. Ethical, legal, and governance issues are only rarely addressed explicitly, with most attention focused on privacy and data protection.

Taken together, these findings suggest that cardiovascular digital twins are technically mature enough to support sophisticated, patient-specific simulations but are not yet ready for routine care. Realizing their potential for precision cardiology will require coordinated progress in standardized evaluation and reporting, rigorous clinical and external validation, user-centered and explainable design, robust data governance, and engagement with regulators and health systems. With the strengthening of these elements, digital twins may evolve from exploratory research tools into trusted, clinically integrated assets for individualized cardiovascular diagnosis, risk assessment, and treatment planning.

Funding

This study was supported by the National Science Foundation with award number 2218046.

XSL•FO
RenderX

Data Availability

All data analyzed in this systematic review were extracted from publicly available publications. The complete list of screened records and their inclusion/exclusion status are provided in Multimedia Appendix 2. The full data extraction matrix for all included studies is provided in Multimedia Appendix 3, and detailed study characteristics (including research question categories, risk-of-bias judgments, and funding information) are summarized in Multimedia Appendix 4.

Authors' Contributions

Conceptualization: FSR, EB, MJ
Data curation: FSR, MJ
Formal analysis: FSR, EB
Methodology: FSR, EB
Supervision: JL
Writing – original draft: FSR, MJ
Writing – review & editing: JL
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File, 29 KB - cardio_v10i1e78499_app1.doc]

Multimedia Appendix 2
Screened records.
[DOCX File, 70 KB - cardio_v10i1e78499_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File, 19 KB - cardio_v10i1e78499_app3.docx]

Multimedia Appendix 4
Research question categories.
[XLSX File, 37 KB - cardio_v10i1e78499_app4.xlsx]

Multimedia Appendix 5
Mapping of raw extraction values to harmonized research question categories.
[XLSX File, 23 KB - cardio_v10i1e78499_app5.xlsx]

Checklist 1
PRISMA checklist.
[PDF File, 213 KB - cardio_v10i1e78499_app6.pdf]

References

1. Grieves M, Vickers J. Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen J, Flumerfelt S, Alves A, editors. Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches: Springer; 2017:85-113. [doi: 10.1007/978-3-319-38756-7_4]
2. Barricelli BR, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access 2019;7:167653-167671. [doi: 10.1109/ACCESS.2019.2953499]
3. Viceconti M, Hunter P. The virtual physiological human: ten years after. Annu Rev Biomed Eng 2016 Jul 11;18:103-123. [doi: 10.1146/annurev-bioeng-110915-114742] [Medline: 27420570]
4. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care: ethical implications of an emerging engineering paradigm. Front Genet 2018;9:31. [doi: 10.3389/fgene.2018.00031] [Medline: 29487613]
5. Sarani Rad F, Hendawi R, Yang X, Li J. Personalized diabetes management with digital twins: a patient-centric knowledge graph approach. J Pers Med 2024 Mar 28;14(4):359. [doi: 10.3390/jpm14040359] [Medline: 38672986]

XSL•FO
RenderX

6. Amini M, Zayeri F, Salehi M. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. BMC Public Health 2021 Feb 25;21(1):401. [doi: 10.1186/s12889-021-10429-0] [Medline: 33632204]
7. Tsampasian V, Bloomfield GS. The evolving global burden of cardiovascular diseases: what lies ahead. Eur J Prev Cardiol 2025 Aug 25;32(11):1016-1017. [doi: 10.1093/eurjpc/zwae330] [Medline: 39365755]
8. Bhagirath P, van der Graaf AWM, de Hooge J, de Groot NMS, Götte MJW. Integrated whole-heart computational workflow for inverse potential mapping and personalized simulations. J Transl Med 2016 May 25;14(1):147. [doi: 10.1186/s12967-016-0902-0] [Medline: 27226006]
9. Seibertz F, Sutanto H, Dülk R, et al. Electrophysiological and calcium-handling development during long-term culture of human-induced pluripotent stem cell-derived cardiomyocytes. Basic Res Cardiol 2023 Apr 5;118(1):14. [doi: 10.1007/s00395-022-00973-0] [Medline: 37020075]
10. Thangaraj PM, Benson SH, Oikonomou EK, Asselbergs FW, Khera R. Cardiovascular care with digital twin technology in the era of generative artificial intelligence. Eur Heart J 2024 Dec 1;45(45):4808-4821. [doi: 10.1093/eurheartj/ehae619] [Medline: 39322420]
11. Plank G, Burton RAB, Hales P, et al. Generation of histo-anatomically representative models of the individual heart: tools and application. Philos Trans A Math Phys Eng Sci 2009 Jun 13;367(1896):2257-2292. [doi: 10.1098/rsta.2009.0056] [Medline: 19414455]
12. Kamel Boulos MN, Zhang P. Digital twins: from personalised medicine to precision public health. J Pers Med 2021 Jul 29;11(8):745. [doi: 10.3390/jpm11080745] [Medline: 34442389]
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
14. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160. [doi: 10.1136/bmj.n160] [Medline: 33781993]
15. Jafarpour M, Bitaraf E, Moeini A, Nahvijou A. Triple A (AAA): a tool to analyze scientific literature metadata with complex network parameters. Presented at: 9th International Conference on Web Research (ICWR); May 3-4, 2023. [doi: 10.1109/ICWR57742.2023.10139229]
16. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
17. Camps J, Wang ZJ, Doste R, et al. Harnessing 12-lead ECG and MRI data to personalise repolarisation profiles in cardiac digital twin models for enhanced virtual drug testing. Med Image Anal 2025 Feb;100:103361. [doi: 10.1016/j.media.2024.103361] [Medline: 39608251]
18. Piersanti R, Bradley R, Ali SY, Quarteroni A, Dede' L, Trayanova NA. Defining myocardial fiber bundle architecture in atrial digital twins. Comput Biol Med 2025 Apr;188:109774. [doi: 10.1016/j.compbiomed.2025.109774] [Medline: 39946790]
19. Torbati S, Daneshmehr A, Pouraliakbar H, et al. Personalized evaluation of the passive myocardium in ischemic cardiomyopathy via computational modeling using Bayesian optimization. Biomech Model Mechanobiol 2024 Oct;23(5):1591-1606. [doi: 10.1007/s10237-024-01856-0] [Medline: 38954283]
20. Yang PC, Jeng MT, Yarov-Yarovoy V, Santana LF, Vorobyov I, Clancy CE. Toward digital twin technology for precision pharmacology. JACC Clin Electrophysiol 2024 Feb;10(2):359-364. [doi: 10.1016/j.jacep.2023.10.024] [Medline: 38069976]
21. Grandits T, Verhulsdonk J, Haase G, Effland A, Pezzuto S. Digital twinning of cardiac electrophysiology models from the surface ECG: a geodesic backpropagation approach. IEEE Trans Biomed Eng 2024 Apr;71(4):1281-1288. [doi: 10.1109/TBME.2023.3331876] [Medline: 38048238]
22. Margara F, Psaras Y, Wang ZJ, et al. Mechanism based therapies enable personalised treatment of hypertrophic cardiomyopathy. Sci Rep 2022 Dec 28;12(1):22501. [doi: 10.1038/s41598-022-26889-2] [Medline: 36577774]
23. Buckler AJ, Marlevi D, Skenteris NT, et al. In silico model of atherosclerosis with individual patient calibration to enable precision medicine for cardiovascular disease. Comput Biol Med 2023 Jan;152:106364. [doi: 10.1016/j.compbiomed.2022.106364] [Medline: 36525832]
24. Zhang Q, Zhang Y, Hao L, et al. A personalized 0D-1D model of cardiovascular system for the hemodynamic simulation of enhanced external counterpulsation. Comput Methods Programs Biomed 2022 Dec;227:107224. [doi: 10.1016/j.cmpb.2022.107224] [Medline: 36379202]
25. Azzolin L, Eichenlaub M, Nagel C, et al. Personalized ablation vs. conventional ablation strategies to terminate atrial fibrillation and prevent recurrence. Europace 2023 Feb 8;25(1):211-222. [doi: 10.1093/europace/euac116] [Medline: 35943361]
26. Jung A, Gsell MAF, Augustin CM, Plank G. An integrated workflow for building digital twins of cardiac electromechanics-a multi-fidelity approach for personalising active mechanics. Mathematics (Basel) 2022 Mar 1;10(5):823. [doi: 10.3390/math10050823] [Medline: 35295404]
27. Gillette K, Gsell MAF, Bouyssier J, et al. Automated framework for the inclusion of a His-Purkinje system in cardiac digital twins of ventricular electrophysiology. Ann Biomed Eng 2021 Dec;49(12):3143-3153. [doi: 10.1007/s10439-021-02825-9] [Medline: 34431016]

XSL•FO
RenderX

28. Patel N, Wiśniowska B, Jamei M, Polak S. Real patient and its virtual twin: application of quantitative systems toxicology modelling in the cardiac safety assessment of citalopram. AAPS J 2017 Nov 27;20(1):6. [doi: 10.1208/s12248-017-0155-8] [Medline: 29181593]
29. Alimohammadi M, Sherwood JM, Karimpour M, Agu O, Balabani S, Díaz-Zuccarini V. Aortic dissection simulation models for clinical support: fluid-structure interaction vs. rigid wall models. Biomed Eng Online 2015 Apr 15;14(34):34. [doi: 10.1186/s12938-015-0032-6] [Medline: 25881252]
30. Gerach T, Schuler S, Fröhlich J, et al. Electro-mechanical whole-heart digital twins: a fully coupled multi-physics approach. Mathematics 2021;9(11):1247. [doi: 10.3390/math9111247]
31. Ooida J, Kiyohara N, Noguchi H, et al. An in silico model for predicting the efficacy of edge-to-edge repair for mitral regurgitation. J Biomech Eng 2024 Feb 1;146(2):021004. [doi: 10.1115/1.4064055] [Medline: 37978048]
32. Margara F, Rodriguez B, Toepfer CN, Bueno-Orovio A. Mavacamten efficacy in mutation-specific hypertrophic cardiomyopathy: an in silico approach to inform precision medicine. Presented at: 2021 Computing in Cardiology (CinC); Sep 13-15, 2021. [doi: 10.23919/CinC53138.2021.9662736]
33. Okada JI, Washio T, Nakagawa M, et al. Multi-scale, tailor-made heart simulation can predict the effect of cardiac resynchronization therapy. J Mol Cell Cardiol 2017 Jul;108(17–23):17-23. [doi: 10.1016/j.yjmcc.2017.05.006] [Medline: 28502795]
34. Sakata K, Bradley RP, Prakosa A, et al. Optimizing the distribution of ablation lesions to prevent postablation atrial tachycardia: a personalized digital-twin study. JACC Clin Electrophysiol 2024 Nov;10(11):2347-2358. [doi: 10.1016/j.jacep.2024.07.002] [Medline: 39243255]
35. Galappaththige S, Gray RA, Costa CM, Niederer S, Pathmanathan P. Credibility assessment of patient-specific computational modeling using patient-specific cardiac modeling as an exemplar. PLoS Comput Biol 2022 Oct;18(10):e1010541. [doi: 10.1371/journal.pcbi.1010541] [Medline: 36215228]
36. Mazumder O, Mukherjee A, Banerjee S, Khandelwal S, Mandana KM, Sinha A. Personalization of a hemodynamic cardiac digital twin: an echocardiogram based approach. Presented at: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 3-6, 2024. [doi: 10.1109/BIBM62325.2024.10822244]
37. Camps J, Wang ZJ, Doste R, et al. Cardiac digital twin pipeline for virtual therapy evaluation. arXiv. Preprint posted online on Jan 18, 2024. [doi: 10.48550/arXiv.2401.10029]
38. Nguyen UC, Cluitmans M, Heijman J, Lumens J. The digital twin. CARIM Midterm Review 2019-2021 Annex 3 Case Studies.: Maastricht University; 2021 URL: https://www.carimmaastricht.org/media_file/080523reviewannex3.pdf#page=30 [accessed 2025-12-16]
39. Hwang T, Lim B, Kwon OS, et al. Clinical usefulness of digital twin guided virtual amiodarone test in patients with atrial fibrillation ablation. NPJ Digit Med 2024 Oct 23;7(1):297. [doi: 10.1038/s41746-024-01298-z] [Medline: 39443659]
40. Falanga M, Cortesi C, Chiaravalloti A, Monte AD, Tomasi C, Corsi C. A digital twin approach for stroke risk assessment in atrial fibrillation patients. Heliyon 2024 Oct 30;10(20):e39527. [doi: 10.1016/j.heliyon.2024.e39527] [Medline: 39512324]
41. Yang PC, Rose A, DeMarco KR, et al. A multiscale predictive digital twin for neurocardiac modulation. J Physiol 2023 Sep;601(17):3789-3812. [doi: 10.1113/JP284391] [Medline: 37528537]
42. Groth C, Porziani S, Biancolini ME, et al. The medical digital twin assisted by reduced order models and mesh morphing. Presented at: International CAE Conference 2018; Oct 8-9, 2018 URL: https://www.researchgate.net/publication/328355804_The_medical_digital_twin_assisted_by_Reduced_Order_Models_and_Mesh_Morphing [accessed 2025-12-16]
43. Zappon E, Gsell MAF, Gillette K, Plank G. Quantifying variabilities in cardiac digital twin models of the electrocardiogram. arXiv. Preprint posted online on Jul 24, 2024. [doi: 10.48550/arXiv.2407.17146]
44. Yin M, Charon N, Brody R, Lu L, Trayanova N, Maggioni M. A scalable framework for learning the geometry-dependent solution operators of partial differential equations. Nat Comput Sci 2024 Dec;4(12):928-940. [doi: 10.1038/s43588-024-00732-2] [Medline: 39653845]
45. B BV, PS P. Enhancing cardiac health: digital twin technology for real- time monitoring and abnormality detection of human heart. Presented at: 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE); Jun 28-30, 2024. [doi: 10.1109/MLISE62164.2024.10674146]
46. Li L, Camps J, Banerjee A, Beetz M, Rodriguez B, Grau V. Deep computational model for the inference of ventricular activation properties. In: Camara O, Puyol-Antón E, Qin C, et al, editors. Statistical Atlases and Computational Models of the Heart Regular and CMRxMotion Challenge Papers STACOM 2022 Lecture Notes in Computer Science: Springer; 2022, Vol. 13593:369-380. [doi: 10.1007/978-3-031-23443-9_34]
47. Sardari Nia P, Ganushchak Y, Maessen J. Proposal of statistical twin as a transition to full digital twin technology for cardiovascular interventions. Interdiscip Cardiovasc Thorac Surg 2024 Mar 29;38(4):ivae032. [doi: 10.1093/icvts/ivae032] [Medline: 38439576]
48. Rahim M, Lalouani W, Toubal E, Emokpae L. A digital twin-based platform for medical cyber-physical systems. IEEE Access 2024;12:174591-174607. [doi: 10.1109/ACCESS.2024.3502077]
49. Alpysbay N, Kolesnikova K, Chinibayeva T. Integrating deep learning and reinforcement learning into a digital twin architecture for medical predictions. Procedia Comput Sci 2024;251:627-632. [doi: 10.1016/j.procs.2024.11.160]

XSL•FO
RenderX

50. Dervişoğlu H, Ülver B, Yurtoğlu RA, Halepmollası R, Haklıdır M. A comparative study on cloud-based and edge-based digital twin frameworks for prediction of cardiovascular disease. Presented at: 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health; Apr 22-24, 2023. [doi: 10.5220/0011859400003476]
51. Álvarez-Barrientos F, Salinas-Camus M, Pezzuto S, Sahli Costabal F. Probabilistic learning of the Purkinje network from the electrocardiogram. Med Image Anal 2025 Apr;101:103460. [doi: 10.1016/j.media.2025.103460] [Medline: 39884028]
52. Scotto A, Giordano N, Rosati S, Balestra G. Design of a digital twin of the heart for the management of heart failure patients. Stud Health Technol Inform 2024 Aug 22;316:875-876. [doi: 10.3233/SHTI240551] [Medline: 39176932]
53. Salvador M, Strocchi M, Regazzoni F, et al. Whole-heart electromechanical simulations using latent neural ordinary differential equations. NPJ Digit Med 2024 Apr 11;7(1):90. [doi: 10.1038/s41746-024-01084-x] [Medline: 38605089]
54. Banerjee A, Camps J, Zacur E, et al. A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices. Philos Trans A Math Phys Eng Sci 2021 Dec 13;379(2212):20200257. [doi: 10.1098/rsta.2020.0257] [Medline: 34689630]
55. Albaghdadi O. Advancing congestive heart failure management through digital twin integration. OSF. Preprint posted online on Mar 24, 2024. [doi: 10.31219/osf.io/pa53r]
56. Marrone S. What does a heart beat for? A heterogeneous approach for human digital twin construction. Procedia Comput Sci 2024;246:5132-5141. [doi: 10.1016/j.procs.2024.09.603]

Abbreviations

AI: artificial intelligence
CFD: computational fluid dynamics
CT: computed tomography
CVD: cardiovascular disease
DL: deep learning
ECG: electrocardiography
FEM: finite element modeling
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
ML: machine learning
MRI: magnetic resonance imaging
PICO: Population, Intervention, Comparison, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
ROBINS-I: Risk of Bias in Non-Randomized Studies - of Interventions
RQ: research question

Edited by A Coristine; submitted 03.Jun.2025; peer-reviewed by AA Arafat, C Exeter, FA Etindele Sosso; accepted 03.Dec.2025; published 08.Jan.2026.

Please cite as:
Sarani Rad F, Bitaraf E, Jafarpour M, Li J
Technologies, Clinical Applications, and Implementation Barriers of Digital Twins in Precision Cardiology: Systematic Review
JMIR Cardio 2026;10:e78499
URL: https://cardio.jmir.org/2026/1/e78499
doi:10.2196/78499

© Fatemeh Sarani Rad, Ehsan Bitaraf, Maryam Jafarpour, Juan Li. Originally published in JMIR Cardio (https://cardio.jmir.org), 8.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bidirectional Associations Between Autonomic Function, Sleep Quality, and Daily Fatigue and Energy Symptoms in Long COVID: Longitudinal Digital Health Study

Nana Yaw Aboagye*, BEng; Mark R Baker*, PhD; Kenneth Baker*, PhD; Silvia Del Din*, PhD

Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, Newcastle, England, United Kingdom

*all authors contributed equally

Corresponding Author:
Nana Yaw Aboagye, BEng
Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, Newcastle, England, United Kingdom

Abstract

Background: Long COVID is characterized by persistent fatigue, with disrupted autonomic function and sleep disturbances frequently reported. Whether daily symptom severity influences subsequent sleep and autonomic recovery (the reverse temporal direction) remains underexplored.

Objective: This study aimed to examine bidirectional, within-person associations between objectively measured sleep quality, nocturnal heart rate variability (HRV), and daily fatigue and energy levels in individuals reporting post-COVID fatigue symptoms.

Methods: We conducted a longitudinal observational study using continuous wearable monitoring (Fitbit Inspire 3) in 14 individuals with long COVID over a median of 28 days (range 12 - 73 d), yielding 678,057 minute-level observations. Nocturnal HRV parameters (root mean square of successive differences [RMSSD], high-frequency power, low-frequency/high-frequency ratio) and sleep metrics (duration, efficiency, deep sleep, rapid eye movement [REM] sleep) were derived automatically. Fatigue (4-level scale) and energy (5-level scale, 0 - 100) were reported via smartphone app (FatigueSense). Two temporal linking strategies were applied: (1) prospective, linking morning/afternoon symptom reports to the previous night's sleep/HRV; and (2) reverse-direction, linking evening reports to the subsequent night's sleep/HRV. Repeated-measures correlation (rmcorr) accounted for within-person clustering; person-mean–centered (Mundlak) cumulative link mixed models disaggregated within-person from between-person effects.

Results: Prospective within-person rmcorr analyses (n=260 observations, 14 participants) identified 3 associations surviving Bonferroni correction (α=.0031): sleep duration (r=−0.276, 95% CI −0.392 to −0.152; P<.001) and REM sleep (r=−0.215, 95% CI −0.342 to −0.079; P=.002) with next-day fatigue, and sleep efficiency with next-day energy in the counterintuitive negative direction (r=−0.196, 95% CI −0.317 to −0.068; P=.003). HRV metrics showed no significant within-person day-to-day associations with symptoms (HRV-RMSSD → fatigue: r=−0.064, P=.36). Mundlak models revealed that HRV-RMSSD associations were predominantly between-person: individuals with chronically higher HRV reported consistently lower fatigue (β=−1.260, 95% CI −2.231 to −0.290; P=.01) and higher energy (β=0.539, 95% CI 0.056 to 1.022; P=.03). A significant Sleep × Steps interaction (β=0.091 per SD of steps per 1000; P=.02) indicated that the protective association of longer sleep on fatigue was attenuated on more active days, consistent with a postexertional malaise pattern. In an exploratory reverse-direction analysis, evening energy (but not fatigue) showed an association with that night's HRV-RMSSD that was nominally significant but did not persist after Bonferroni correction (rmcorr: r=0.327, 95% CI 0.080 to 0.537; P=.01; α=.0050) and was supported by a log-linear mixed model (β=0.0069; P=.02).

Conclusions: In this exploratory longitudinal study, sleep duration and REM sleep showed day-to-day (within-person) associations with next-day fatigue, while HRV-RMSSD distinguished individuals with better versus worse average symptom burden. The counterintuitive sleep efficiency-energy association most plausibly reflects chance, given the null sleep efficiency-fatigue association and the multiple-testing context. The preliminary, uncorrected reverse-direction association between evening energy and nocturnal HRV-RMSSD warrants replication. Experimental work is needed to test causal mechanisms and clinical relevance.

(JMIR Cardio 2026;10:e99630) doi:10.2196/99630

KEYWORDS

post-COVID fatigue symptoms; post-COVID-19 condition; heart rate variability; sleep quality; fatigue; wearable devices; exploratory longitudinal study

XSL•FO
RenderX

Introduction

Long COVID, Sleep, and Autonomic Dysfunction

Long COVID, or postacute sequelae of SARS-CoV-2 infection, affects many individuals following COVID-19 infection, resulting in persistent and often debilitating symptoms lasting months to years beyond the acute infection phase [1-4]. Among the most commonly reported symptoms are profound fatigue, postexertional malaise (PEM), and sleep disturbances, which collectively impair quality of life and functional capacity [5,6]. Understanding the physiological mechanisms underlying these symptoms and their temporal dynamics is critical for developing effective management strategies.

Sleep disturbances are prevalent in long COVID, with studies in post-COVID cohorts reporting sleep disruption in 40% to 75% of affected individuals [1,6]. Concurrently, disruptions in autonomic nervous system function have been documented, including reduced heart rate variability (HRV, a marker of parasympathetic activity and cardiovascular adaptability) [7,8]. Lower HRV has been associated with worse long COVID symptoms, suggesting impaired autonomic regulation may contribute to symptom persistence [9,10].

HRV, particularly during sleep when parasympathetic activity normally predominates, provides a window into autonomic recovery processes [11]. The high-frequency component of HRV (HRV-HF, 0.15 - 0.4 Hz) reflects vagal/parasympathetic modulation, while the low-frequency component (HRV-LF, 0.04 - 0.15 Hz) reflects mixed sympathetic and parasympathetic influences [12]. The LF/HF ratio is often interpreted as an index of sympathovagal balance, though its physiological specificity is debated [13]. In healthy individuals, sleep is characterized by stage-dependent autonomic modulation, with parasympathetic predominance during non–rapid eye movement (REM) sleep and relative sympathetic activation during REM [14].

From Prediction to Bidirectional Association

Previous research has focused on unidirectional relationships, investigating how sleep quality or HRV predicts subsequent symptom severity [15]. A critical gap exists in understanding the reverse direction: whether daily symptom severity influences physiological recovery during subsequent sleep. This bidirectional perspective is particularly relevant in long COVID, where symptom-driven behavioral adaptations and psychological states may impact sleep quality and autonomic recovery [16].

The Role of Physical Activity

PEM, a hallmark of long COVID and myalgic encephalomyelitis/chronic fatigue syndrome, involves symptom worsening following physical or cognitive exertion [17-19]. COVID-era physical inactivity has been documented as an additional concern for individuals recovering from SARS-CoV-2 infection [20]. The relationships between daily activity levels, sleep quality, and symptoms remain poorly understood. Evidence from accelerometer-based studies suggests that sleep duration, sedentary behavior, and physical activity should be considered as codependent daily behaviors rather than isolated exposures [21]. Activity may act as a moderator, with effects on symptoms depending on sleep quality.

Study Aim

To address these gaps, we conducted an exploratory longitudinal study using continuous wearable device monitoring in individuals reporting post-COVID fatigue. We employed a bidirectional analytical approach examining (1) prospective associations: how does the previous night's sleep/HRV relate to next-day symptoms?; (2) reverse-direction associations: how do evening symptoms relate to that night's sleep/HRV?; and (3) activity moderation: does physical activity interact with sleep quality in predicting symptoms? We hypothesized that (H1) sleep duration would show within-person day-to-day associations with symptoms; (H2) evening symptom severity would predict subsequent nocturnal autonomic function; and (H3) physical activity would moderate the sleep-symptom relationship.

Methods

Study Design and Participants

This was a longitudinal observational study conducted between May 2025 and September 2025. Participants were recruited from individuals with post-COVID fatigue who had previously participated in a National Institute for Health and Care Research funded study at Newcastle University (Pausing-PCF [22]) and were engaged in patient and public involvement and engagement activities, as well as through online patient communities and social media. Inclusion was based on self-reported post-COVID fatigue symptoms and willingness to wear the device continuously. No formal clinician-adjudicated diagnosis was required, and no healthy control group was recruited; this represents an acknowledged limitation (see *Limitations*). This study uses a different cohort, device, and analytical framework from the authors' prior feasibility study in this population [15], which involved 68 different participants, a different wearable device (Axivity AX6 accelerometer with VitalPatch ECG patch), and did not include reverse-direction associations or within-person effect disaggregation.

Ethical Considerations

The study protocol was approved by the Newcastle University Ethics Committee (reference 47832 - 2023). All procedures were conducted in accordance with relevant guidelines and regulations, and all participants provided written informed consent prior to enrollment. Participant data were deidentified prior to analysis. No individual identifying information appears in any figures or tables. Participants were volunteers who received no financial compensation. The ethics approval permitted secondary analysis of wearable data collected under this protocol without additional consent.

Wearable Device Monitoring

Participants wore a Fitbit Inspire 3 wristband continuously (24 h/d) for the study duration, charging for approximately 1 to 2 hours per day during inactive periods (eg, while showering). The Inspire 3 uses photoplethysmography to estimate interbeat intervals, from which HRV metrics are derived using proprietary algorithms. It is important to note that validation studies of consumer wearable photoplethysmography sensors have primarily examined heart rate accuracy and sleep staging

XSL•FO
RenderX

[23-25]. Where spectral HRV has been formally evaluated in ring- and wrist-worn photoplethysmography devices against electrocardiogram (ECG) during nocturnal monitoring, agreement has been inconsistent: nightly-average HF power shows moderate-to-good correspondence, but 5-minute segment estimates of LF power (r≈0.42) and LF/HF ratio (r≈0.36) show poor agreement with ECG-derived values [26,27]. No independent peer-reviewed study has specifically validated the Fitbit Inspire 3's proprietary spectral HRV algorithm against ECG; these metrics should therefore be treated as indicative rather than definitive measures of autonomic function. The device's internal HRV sampling algorithm processes photoplethysmography segments during sleep, enabling frequency-domain HRV estimation [28]; the "1 Hz" sampling figure sometimes quoted for this device refers to display resolution rather than the raw optical sampling rate, which is substantially higher. Participants were instructed that step data would be collected automatically during waking hours.

Sleep metrics extracted included sleep duration (min), sleep efficiency (proportion of time asleep while in bed, %), deep sleep duration (min), and REM sleep duration (min). The device automatically detected sleep onset, awakening, and sleep stages using movement patterns and heart rate data [29].

HRV metrics included HRV-RMSSD (root mean square of successive differences), a time-domain measure reflecting parasympathetic activity [30]; HRV-HF (high-frequency power, 0.15 - 0.4 Hz), a frequency-domain measure of vagal modulation [12]; HRV-LF (low-frequency power, 0.04 - 0.15 Hz); and the LF/HF ratio, included as an exploratory autonomic index but interpreted cautiously given its debated physiological specificity [13]. HRV metrics were measured exclusively during sleep, providing an index of nocturnal autonomic recovery. Sleep periods with <50% HRV coverage were excluded, as signal quality is lowest during restless sleep. Physical activity was quantified as daily step count accumulated between sleep offset and symptom report time (prospective analysis) or from sleep offset to evening report (reverse-direction analysis).

Symptom Reporting

Participants completed brief symptom reports via the FatigueSense smartphone app [31] at their convenience throughout each day. Reports were automatically timestamped. Participants rated fatigue severity on a 4-level categorical scale (none, mild, moderate, severe); and energy level on a 5-level ordinal scale (0, 25, 50, 75, 100). Reports were categorized by time of day: morning (5 AM-11:59 AM); afternoon (noon-4:59 PM); evening (5 PM-9:59 PM); and night (10 PM-4:59 AM). Participants were encouraged to report at least once daily but were not assigned fixed reporting times, introducing potential variability in the gap between waking and morning report completion and between evening report and subsequent sleep onset.

Temporal Linking Strategy

A key methodological feature of this study was the explicit use of symptom report timestamps to enable bidirectional analysis. Two analytically distinct datasets were created.

Prospective Linking (Sleep/HRV → Symptoms)

Morning and afternoon symptom reports were linked to the previous night's sleep and HRV data. For each report, we extracted sleep metrics from the most recent completed sleep period, HRV metrics averaged across that sleep period, and step count accumulated between sleep offset and symptom report time. This addresses the question: "How does last night's sleep and autonomic activity predict how I feel today?"

Reverse-Direction Linking (Symptoms → Sleep/HRV)

Evening symptom reports were linked to the subsequent night's sleep/HRV (the sleep period beginning after the evening report). Step count for this dataset was accumulated from previous sleep offset to evening report time. This addresses the question: "How is tonight's sleep and autonomic recovery associated with how I felt this evening?" Note that this direction is temporally prospective (symptoms precede the sleep period); we use "reverse-direction" to distinguish it from the prospective symptom-prediction analysis.

Data Preprocessing

Data were extracted via the Fitbit Web API. HRV metrics (RMSSD, HF power, and LF power) were obtained from the intraday HRV end point (fields hrvRmssd, hrvHf, and hrvLf), which returns values at a native 5-minute resolution during device-detected sleep; the device does not expose beat-to-beat interbeat intervals through this endpoint. For each sleep period, HRV metrics were averaged across all available 5-minute segments, and periods with <50% segment coverage were excluded. Across the cohort, 36,521 five-minute HRV segments were recorded over 471 sleep periods (mean ≈78 segments, ≈6.5 h per period); thus, the "HRV segments" we report denote counts of 5-minute windows rather than individual minute-level readings. Sleep periods were identified using device-detected onset and offset times. Daily aggregates were computed for each 24-hour period (noon to 11:59 AM the following day). Symptom reports were matched to the appropriate sleep period based on timestamps.

Statistical Analysis

All analyses were conducted in Python 3.13 and R (version 4.5.2; R Foundation for Statistical Computing) using the ordinal, lme4, lmerTest, repeated-measures correlation (rmcorr), and car packages.

Descriptive Statistics

Summary statistics were computed for all variables. All continuous variables showed significant departures from normality (Shapiro-Wilk test, P<.05), informing the choice of methods. Fatigue was a 4-level ordinal scale; no participants reported "None," so 3 categories were observed (mild, moderate, severe). Energy was a 5-level ordinal scale; participants rarely selected 0 or 100. Restricted variance in both scales represents a measurement limitation (see *Limitations*).

Observation Flow and Sample Sizes

Of 260 prospective observations linked to a prior sleep period, 222 had complete HRV data (RMSSD+LF/HF). Complete-case counts differed by one observation between outcomes because

XSL•FO
RenderX

of a single missing energy rating: the parsimonious cumulative link mixed models (CLMMs) used 215 (fatigue) and 214 (energy) observations, and the comprehensive and Mundlak CLMMs used 211 (fatigue) and 210 (energy). Of 105 reverse-direction observations, 71 had complete HRV data for the reverse-direction model. A total of (n=243, 93.5%) prospective observations had step data. Intraclass correlation coefficients (ICCs) were estimated from null CLMMs: fatigue ICC=0.245 (24.5% between-person variance); energy ICC=0.491 (49.1% between-person variance). The high ICC for energy indicates that between-person differences are substantial, necessitating disaggregation of within- and between-person effects.

Bivariate Associations: rmcorr

Bivariate associations were assessed using rmcorr [32], which accounts for within-person clustering by partialling out participant-level mean differences. This approach avoids the pseudoreplication error that occurs when pooling repeated observations from the same individuals and treating them as independent. Pooled Spearman correlations are presented in Multimedia Appendix 1 for comparison, but rmcorr estimates are the primary bivariate results. Statistical significance was assessed at α=.05 with Bonferroni correction within each domain (prospective: 16 tests, $\alpha_{corrected}$=.0031; reverse-direction: 10 tests, $\alpha_{corrected}$=.0050).

Mixed-Effects Models

Prospective Models (Mundlak Specification)

Given the substantial between-person variance, particularly for energy (ICC=0.491), we employed person-mean–centered (Mundlak) CLMMs to disaggregate within-person (day-to-day) from between-person effects [33]. For each predictor X, we computed the person mean (between-person effect, hereafter "bp_") and the deviation from the person mean (within-person effect, hereafter "wp_"). Both were included as separate predictors in the CLMM. This allows the within-person coefficient (the primary scientific question: "Does your HRV tonight predict your symptoms tomorrow compared to your own average?") to be estimated independently of stable between-person differences.

Standard CLMMs (without mean-centering) were also fitted as a sensitivity analysis, using parsimonious (sleep duration+HRV-RMSSD+LF/HF ratio) and comprehensive (all sleep metrics+HRV metrics) specifications. Random intercept-only models were used throughout; likelihood ratio tests confirmed that random slopes did not improve fit (fatigue: likelihood ratio P=.52; energy: likelihood ratio P=.87).

Variance inflation factors (VIFs) were computed for the comprehensive model to assess collinearity. All VIFs were <2 (range 1.10 - 1.84), indicating no meaningful multicollinearity.

All continuous predictors were z score standardized (mean 0, SD 1) to enable effect size comparison via standardized coefficients. 95% CIs are reported for all model estimates.

Reverse-Direction Model

A linear mixed-effects model (lmer) was fitted predicting log-transformed nocturnal HRV-RMSSD from evening fatigue and energy, with random intercepts for participants. Log transformation was applied because HRV-RMSSD is right-skewed; Shapiro-Wilk tests confirmed substantially improved residual normality under log transformation (raw: W=0.646, P<.001; log: W=0.941, P=.002), although some departure remained, which is noted as a limitation. The coefficient for energy is back-transformed to a percentage change per 25-point energy increase. We note that energy is modeled as an ordinal outcome in the CLMMs but enters this reverse-direction model as a continuous predictor of HRV-RMSSD. This continuous treatment is a deliberate, justified simplification: the 5 energy levels are anchored at equally spaced numeric values (0, 25, 50, 75, 100), so treating them as interval-scaled for use as a predictor is reasonable and yields the interpretable "per 25-point increase" effect; the ordinal treatment is retained where energy is the outcome because the proportional-odds model makes no equal-spacing assumption. The 2 treatments are therefore applied consistently to each role (energy-as-outcome vs energy-as-predictor) rather than inconsistently to the same quantity.

Activity Pattern Analysis

Activity Tertiles

Participants were stratified into low, medium, and high activity groups based on step count tertiles. Mean fatigue and energy were compared across groups using Kruskal-Wallis tests with pairwise comparisons.

PEM and Sleep × Steps Interaction

To formally test whether physical activity moderated the sleep-symptom relationship, a linear mixed-effects model was fitted predicting fatigue (numeric) from sleep duration (z scored), steps per 1000 (z scored), and their interaction (sleep duration × steps/1000). Steps were rescaled to per-1000 units to yield interpretable coefficients. This continuous interaction model tests whether the association between longer sleep and lower fatigue differs depending on activity level.

Results

Sample Characteristics

Fourteen individuals reporting post-COVID fatigue symptoms (12 female [85.7%], 2 male [14.3%]; age range 35 - 64 y) enrolled and completed the study protocol. Age distribution: 35 to 44 years (n=4, 28.6%), 45 - 54 years (n=3, 21.4%), 55 - 64 years (n=7, 50%). Participant characteristics and monitoring details are summarized in Table 1. Descriptive statistics for all variables are presented in Table 2.

XSL•FO
RenderX

Table . Participant and study characteristics (N=14).

Characteristic	Value
Sex, n (%)	
Female	12 (85.7)
Male	2 (14.3)
Age range (y)	35 - 64
35 - 44, n (%)	4 (28.6)
45 - 54, n (%)	3 (21.4)
55 - 64, n (%)	7 (50)
Monitoring duration	
Median days worn (range)	28 (12 - 73)
Total minute-level observations	678,057
Total sleep periods	471
Total HRV[a] 5-minute segments recorded	36,521
Mean HRV coverage per sleep period, % (SD)	77.6 (29.4)
Symptom reports (N=598), n (%)	
Morning reports (5 AM-11:59 AM)	261 (43.6)
Afternoon reports (noon-4:59 PM)	83 (13.9)
Evening reports (5 PM-9:59 PM)	204 (34.1)
Night reports (10 PM-4:59 AM)	50 (8.4)
Analytical datasets	
Prospective observations, raw (with complete HRV)	260 (222)
Complete data: parsimonious CLMM[b] (fatigue/energy)	215/214
Complete data: comprehensive and Mundlak CLMM (fatigue/energy)	211/210
Observations with step data, n (%)	243 (93.5)
Participants in prospective analysis	14
Reverse-direction observations, raw (with complete HRV)	105 (71)
Participants in reverse-direction analysis	12
Fatigue ICC[c] (null CLMM)	0.245
Energy ICC (null CLMM)	0.491

[a]HRV: heart rate variability.

[b]CLMM: cumulative link mixed models.

[c]ICC: intraclass correlation coefficients.

XSL•FO
RenderX

Table . Descriptive statistics for sleep, heart rate variability, activity, and symptom variables[a].

Variable	Values
Sleep metrics, mean (SD)	
Sleep duration, min	470 (103)
Sleep efficiency, %	90.2 (9.9)
Deep sleep, min	65 (26)
REM[b] sleep, min	84 (33)
HRV[c] metrics (nocturnal), mean (SD)	
HRV-RMSSD[d], ms	34.0 (23.3)
HRV-HF[e], ms^2	495 (1337)
HRV-LF[f], ms^2	1181 (2057)
LF/HF ratio	4.51 (3.08)
Physical activity, mean (SD)	
Daily steps (waking window)	7593 (3806)
Symptoms, n (%)	
Fatigue	
Mild	170 (28.5)
Moderate	388 (65.0)
Severe	39 (6.5)
Energy	
25	141 (23.7)
50	336 (56.4)
75	117 (19.6)
100	2 (0.3)

[a]HRV metrics were measured during sleep only. Daily steps reflect waking-window accumulation between sleep offset and symptom report time. All continuous variables showed significant departures from normality (Shapiro-Wilk test, $P<.05$). Sleep, HRV and activity statistics are computed over the 260 linked prospective observations; symptom counts are over all reports. Symptom percentages are of reports with a valid rating for that item (fatigue n=597; energy n=596, of 598 total reports).

[b]REM: rapid eye movement.

[c]HRV: heart rate variability.

[d]RMSSD: root mean square of successive differences.

[e]HF: high-frequency.

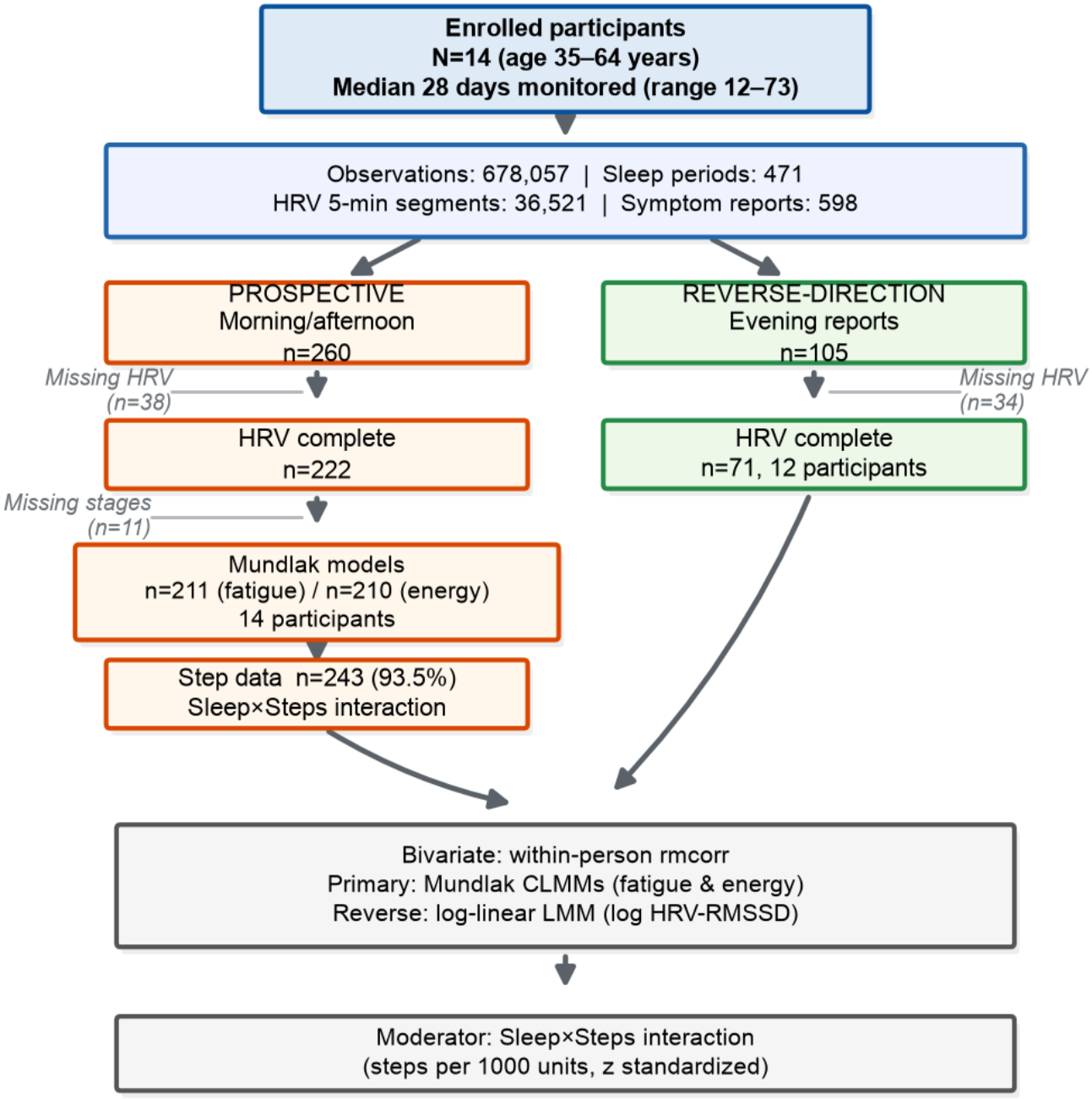
[f]LF: low-frequency.

Prospective Associations: Sleep and HRV Predicting Next-Day Symptoms

We examined how the previous night's sleep and HRV were associated with next-day fatigue and energy (n=260 prospective observations, 14 participants). Figure 1 illustrates the observation flow.

XSL•FO
RenderX

Figure 1. Observation and analysis flow diagram. CLMM: cumulative link mixed model; HRV: heart rate variability; LMM: linear mixed model; rmcorr: repeated-measures correlation; RMSSD: root mean square of successive differences.



Within-Person Bivariate Associations (rmcorr)

Within-person rmcorrs are presented in Table 3 and Figure 2.

XSL•FO
RenderX

Table . Prospective within-person associations: repeated-measures correlations[a].

Predictor	Fatigue, *r* (95% CI)	*P* value	Energy, *r* (95% CI)	*P* value
Sleep duration	−0.276 (−0.392 to −0.152)	<.001[b]	0.176 (0.048 to 0.299)	.007
Sleep efficiency	−0.008 (−0.137 to 0.122)	.91	−0.196 (−0.317 to −0.068)	.003[b]
Deep sleep	−0.127 (−0.261 to 0.012)	.07	0.138 (−0.001 to 0.272)	.05
REM[c] sleep	−0.215 (−0.342 to −0.079)	.002[b]	0.064 (−0.075 to 0.200)	.37
HRV[d]-RMSSD[e]	−0.064 (−0.198 to 0.072)	.36	0.016 (−0.120 to 0.152)	.81
HRV-HF[f]	−0.064 (−0.198 to 0.072)	.36	0.020 (−0.116 to 0.156)	.77
HRV-LF[g]	−0.120 (−0.252 to 0.016)	.08	0.047 (−0.090 to 0.182)	.50
LF/HF ratio	−0.044 (−0.179 to 0.092)	.53	0.063 (−0.074 to 0.197)	.37

[a]Bonferroni-corrected threshold for 16 tests: α=.0031.

[b]Survives Bonferroni correction (*P*<.003). Correlations are within-person (day-to-day) estimates using the rmcorr package; they reflect associations net of stable between-person differences. Analytic N varies by predictor owing to missing data (fatigue: 214 - 243; energy: 213 - 242); participants n=14.

[c]REM: rapid eye movement.

[d]HRV: heart rate variability.

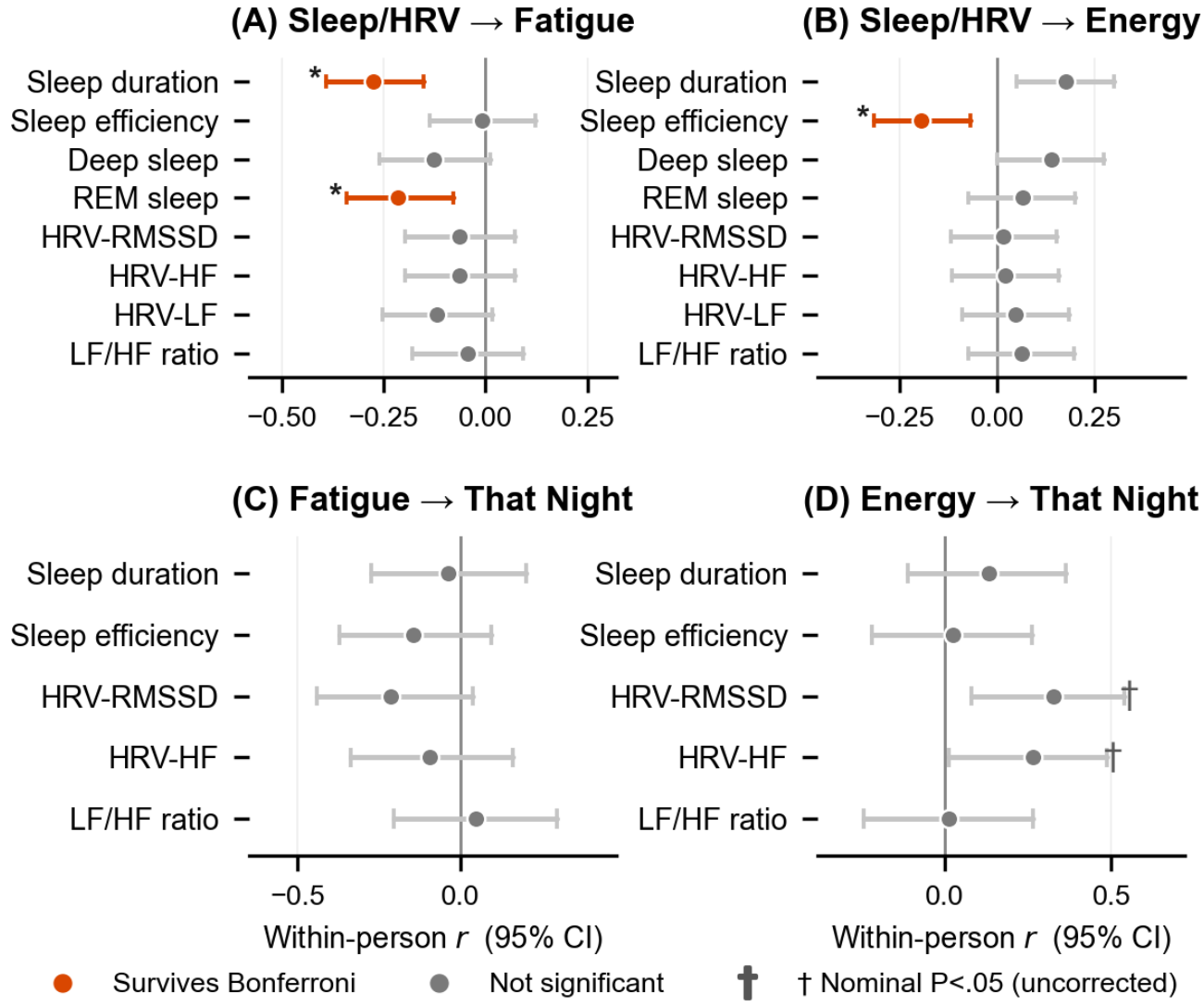
[e]RMSSD: root mean square of successive differences.

[f]HF: high-frequency.

[g]LF: low-frequency.

XSL•FO
RenderX

Figure 2. Bidirectional within-person associations (repeated-measures correlation), with points representing within-person *r* with their 95% CI. CLMM: cumulative link mixed model; HF: high-frequency; HRV: heart rate variability; LF: low-frequency; REM: rapid eye movement; RMSSD: root mean square of successive differences.



After Bonferroni correction (α=.0031, 16 tests), 3 within-person associations survived: sleep duration was negatively associated with next-day fatigue (r=−0.276, 95% CI −0.392 to −0.152; P<.001); REM sleep was negatively associated with fatigue (r=−0.215, 95% CI −0.342 to −0.079; P=.002); and sleep efficiency was negatively associated with energy (r=−0.196, 95% CI −0.317 to −0.068; P=.003). The direction of this last association was counterintuitive, in that higher sleep efficiency was associated with lower next-day energy. Given that sleep efficiency showed no association with next-day fatigue (r=−0.008, P=.91) and the number of tests conducted, it is most plausibly a chance finding and requires replication. Sleep duration was also positively associated with next-day energy (r=0.176, 95% CI 0.048 to 0.299; P=.007), and deep sleep showed marginal associations with both fatigue (r=−0.127, P=.07) and energy (r=0.138, P=.05); none of these survived Bonferroni correction.

Notably, no HRV metric showed a significant within-person day-to-day association with next-day fatigue or energy after correcting for multiple comparisons (HRV-RMSSD → fatigue: r=−0.064, 95% CI −0.198 to 0.072; P=.36; HRV-RMSSD → energy: r=0.016, 95% CI −0.120 to 0.152; P=.81; HRV-HF → fatigue: r=−0.064, P=.36; HRV-HF → energy: r=0.020, P=.77). These within-person results contrast with pooled Spearman correlations (presented in Multimedia Appendix 1), which showed associations of ρ=−0.317 (HRV-RMSSD → fatigue) and ρ=0.416 (HRV-RMSSD → energy). The discrepancy reflects inflation of pooled correlations by stable between-person differences in HRV and symptom burden (ICC_energy=0.491), and underscores the importance of within-person analysis for repeated-measures data.

Person-Mean–Centered Mixed-Effects Models (Mundlak)

Mundlak CLMMs disaggregating within-person from between-person effects are presented in Table 4 and visualized in Figure 3.

XSL•FO
RenderX

Table . Person-mean–centered (Mundlak) cumulative link mixed models: within-person and between-person effects on fatigue and energy[a].

Effect	Fatigue, β (95% CI)	*P* value	Energy, β (95% CI)	*P* value
Within-person (wp_)[b]				
wp_sleep duration	−0.610 (−0.937 to −0.283)	<.001	0.368 (0.066 to 0.671)	.02
wp_HRV[c]-RMSSD[d]	−0.130 (−0.437 to 0.177)	.41	0.144 (−0.166 to 0.453)	.36
wp_LF[e]/HF[f]	−0.215 (−0.523 to 0.093)	.17	0.135 (−0.158 to 0.427)	.37
Between-person (bp_)[g]				
bp_sleep duration	−0.509 (−1.101 to 0.083)	.09	0.220 (−0.517 to 0.958)	.56
bp_HRV-RMSSD	−1.260 (−2.231 to −0.290)	.01	0.539 (0.056 to 1.022)	.03
bp_LF/HF	0.313 (−0.267 to 0.893)	.29	−0.746 (−1.637 to 0.145)	.10

[a]All predictors *z* score standardized (mean 0, SD 1). Random intercepts for participants. intraclass correlation coefficient_fatigue=0.245; intraclass correlation coefficient_energy=0.491.

[b]wp_ : within-person deviation from own mean.

[c]HRV: heart rate variability.

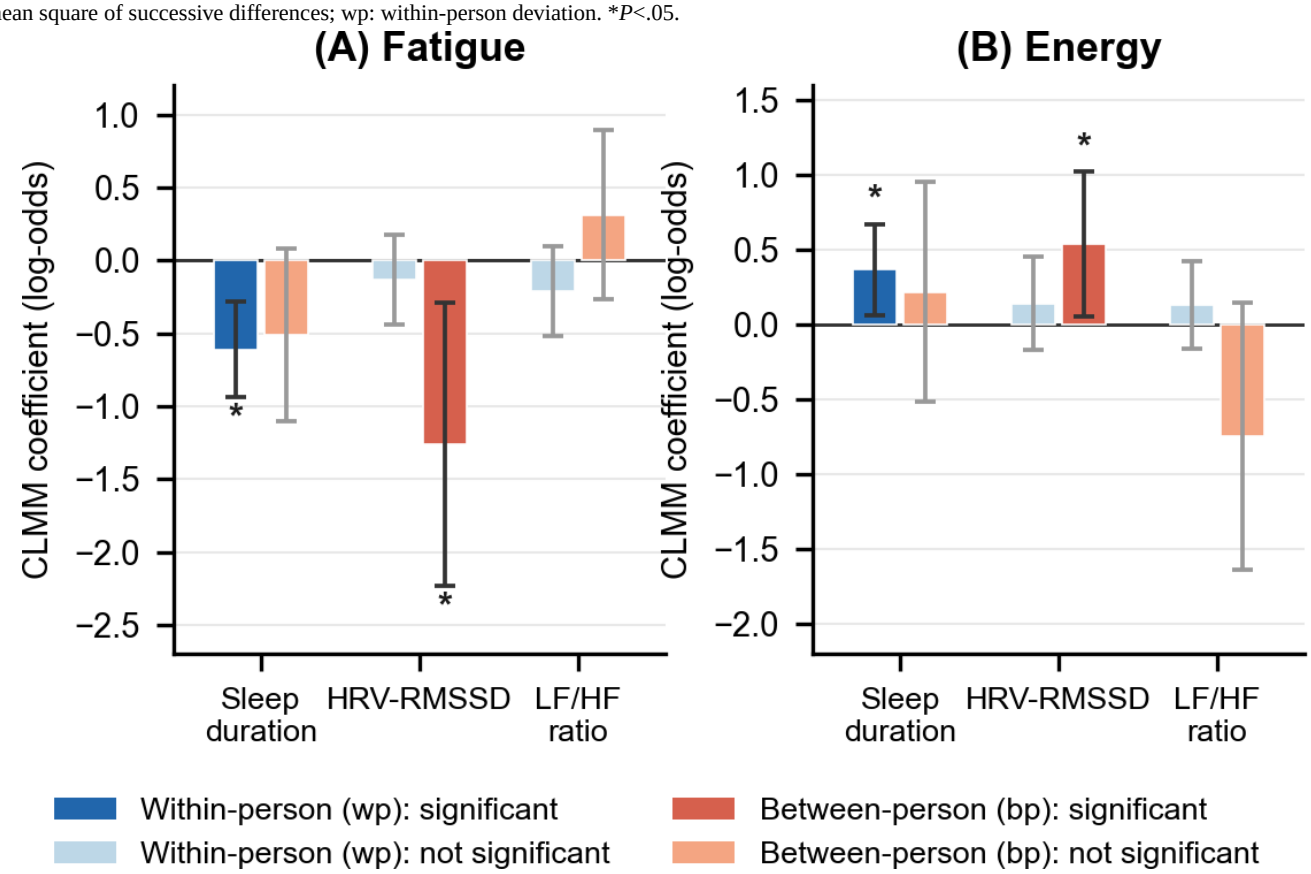
[d]RMSSD: root mean square of successive differences.

[e]LF: low-frequency.

[f]HF: high-frequency.

[g]bp_ : between-person mean.

Figure 3. Within-person versus between-person effects (Mundlak cumulative link mixed models; n=211 [fatigue]/n=210 [energy], 14 participants). bp: between-person mean; CLMM: cumulative link mixed model; HF: high-frequency; HRV: heart rate variability; LF: low-frequency; RMSSD: root mean square of successive differences; wp: within-person deviation. *P<.05.



The analyses included 211 observations for fatigue and 210 observations for energy from 14 participants. The energy model had 1 fewer complete case than the fatigue model, consistent with the parsimonious and comprehensive CLMMs (Tables S1 and S2 in Multimedia Appendix 1).

For fatigue, within-person (day-to-day) sleep duration was a significant predictor (wp_sleep duration: β=−0.610, 95% CI −0.937 to −0.283; *P*<.001), indicating that on nights with longer sleep than a participant's own average, next-day fatigue tended to be lower. The within-person HRV-RMSSD effect was not

XSL•FO
RenderX

significant (wp_hrv_rmssd: β=−0.130, 95% CI −0.437 to 0.177; *P*=.41). However, the between-person HRV-RMSSD effect was significant (bp_hrv_rmssd: β=−1.260, 95% CI −2.231 to −0.290; *P*=.01), indicating that participants with chronically higher mean HRV-RMSSD reported consistently lower fatigue.

For energy, within-person sleep duration was again a significant predictor (wp_sleep duration: β=0.368, 95% CI 0.066 to 0.671; *P*=.02). The within-person HRV-RMSSD effect was not significant (wp_hrv_rmssd: β=0.144, 95% CI −0.166 to 0.453; *P*=.36), while the between-person effect was significant (bp_hrv_rmssd: β=0.539, 95% CI 0.056 to 1.022; *P*=.03).

VIFs for all predictors ranged from 1.10 to 1.84, confirming no problematic collinearity.

Sensitivity Analyses: Standard CLMMs

For comparison, standard CLMMs are presented in Multimedia Appendix 1 (Tables S1 and S2). In the parsimonious model (n=215 for fatigue, n=214 for energy), sleep duration (fatigue: β=−0.702, *P*<.001; energy: β=0.451, *P*=.006) and HRV-RMSSD (fatigue: β=−0.492, *P*=.05; energy: β=0.613, *P*=.009) were both significant. In the comprehensive model (n=211), only HRV-RMSSD remained significant (fatigue: β=−0.705, 95% CI −1.324 to −0.086, *P*=.03; energy: β=0.546, 95% CI 0.104 to 0.988, *P*=.02). These results should be interpreted with caution because the standard CLMMs do not disaggregate within-person from between-person effects, and with energy ICC=0.491 the estimates reflect a mix of both. Primary analysis was therefore based on the Mundlak models.

Reverse-Direction Associations: Evening Symptoms Predicting Nocturnal HRV

We examined whether evening symptom severity predicted that night's sleep and HRV (n=105 reverse-direction observations, 71 with complete HRV, 12 participants). These results are presented in Table 5.

Table . Reverse-direction within-person associations: evening symptoms predicting same-night sleep/heart rate variability[a].

Outcome	Fatigue, *r* (95% CI)	*P* value	Energy, *r* (95% CI)	*P* value
Sleep duration	−0.037 (−0.273 to 0.203)	.77	0.134 (−0.110 to 0.362)	.28
Sleep efficiency	−0.146 (−0.371 to 0.096)	.24	0.024 (−0.217 to 0.263)	.85
HRV[b]-RMSSD[c]	−0.215 (−0.443 to 0.039)	.10	0.327 (0.080 to 0.537)	.01[d]
HRV-HF[e]	−0.094 (−0.337 to 0.162)	.47	0.265 (0.012 to 0.486)	.04[d]
LF[f]/HF ratio	0.048 (−0.206 to 0.297)	.71	0.013 (−0.242 to 0.266)	.92

[a]Bonferroni-corrected threshold for 10 tests: α=.0050.

[b]HRV: heart rate variability.

[c]RMSSD: root mean square of successive differences.

[d]Nominally significant (*P*<.05) but does not survive Bonferroni correction (α=.0050); no reverse-direction bivariate association survived correction. The evening energy → HRV-RMSSD association is independently supported by the log-linear mixed model (β=0.0069, *P*=.02). Analytic N varies by outcome owing to missing data (fatigue: 72 - 81; energy: 71 - 80); HRV outcomes include 12 participants, sleep outcomes 14. rmcorr within-person estimates.

[e]HF: high-frequency.

[f]LF: low-frequency.

Within-Person Bivariate Associations (rmcorr)

Higher evening energy showed a nominally significant positive within-person association with that night's HRV-RMSSD (*r*=0.327, 95% CI 0.080 to 0.537; *P*=.01), though this did not survive Bonferroni correction (α_corrected=.0050, 10 tests). Neither HRV-HF (*r*=0.265, *P*=.04) nor any other reverse-direction bivariate association survived correction. Evening fatigue showed no significant associations with any sleep or HRV outcome (HRV-RMSSD: *r*=−0.215, *P*=.10; all others *P*>.20). Neither symptom predicted sleep duration or efficiency.

Log-Linear Mixed Model

A log-linear mixed-effects model (predicting log [HRV-RMSSD] from evening fatigue and energy) was consistent with the within-person association: evening energy independently predicted that night's log-transformed HRV-RMSSD (β=0.0069, 95% CI 0.0013 to 0.0125; *P*=.02), while fatigue did not (β=−0.029, *P*=.69). Each 25-point increase in evening energy (on the 0 - 100 scale) was associated with approximately 18.8% higher nocturnal HRV-RMSSD (ratio: 1.188, corresponding to approximately 6.4 ms at the mean level of 34.0 ms). As this association did not survive Bonferroni correction in the bivariate analysis, it is interpreted as exploratory (see *Discussion*).

Residual normality under log transformation (Shapiro-Wilk *W*=0.941, *P*=.002) was substantially better than the raw model (*W*=0.646, *P*<.001), supporting the log-transformed approach, although some departure from normality remained.

Physical Activity Patterns and Symptom Interactions

Activity Tertiles

Activity tertile thresholds were low (<5684 steps), medium (5684 - 8545 steps), and high (>8545 steps). Mean fatigue and energy levels showed minimal differences across tertiles (low: fatigue=1.83, energy=51; medium: fatigue=1.78, energy=47; high: fatigue=1.72, energy =50). Kruskal-Wallis tests confirmed

XSL•FO
RenderX

these differences were not statistically significant (fatigue: χ^2_2=1.102, *P*=.58; energy: χ^2_2=1.879, *P*=.39).

Sleep × Steps Interaction

To formally test whether activity moderated the sleep-symptom relationship, a linear mixed-effects interaction model was fitted. Sleep duration (*z* scored) and steps per 1000 (*z* scored) were included as main effects along with their interaction:

- fatigue_num ~ sleep duration_*z* × steps_per 1000_*z* + (1 | participant)

The Sleep × Steps interaction was significant (β=0.091, 95% CI 0.018 to 0.164; *P*=.02). The positive coefficient indicates that the protective association of longer sleep on fatigue was weaker (attenuated) at higher activity levels. This pattern is consistent with a PEM mechanism: on more active days, the symptom-protective effect of sleep may be partially offset by activity-related physiological burden. Step count alone was not significantly associated with fatigue (β=−0.056, *P*=.09) or energy (tertile analysis, *P*>.39), indicating no simple linear activity-symptom relationship.

Step data were available for 93.5% (243/260) of prospective observations. This figure represents the proportion of waking windows between sleep offset and symptom report in which at least some step activity was recorded. This differs from the proportion of all monitored waking-period time with continuous step activity, which was lower due to sedentary periods.

Discussion

Summary of Main Findings

This exploratory longitudinal wearable study examined bidirectional associations between sleep, nocturnal autonomic function, and daily symptoms in individuals with long COVID. The main findings were as follows: (1) sleep duration and REM sleep showed significant within-person (day-to-day) associations with next-day fatigue, and sleep efficiency showed a counterintuitive negative association with next-day energy that is most plausibly a chance finding; (2) HRV-RMSSD was a significant between-person predictor; individuals with chronically higher HRV reported consistently lower fatigue and higher energy but did not show significant within-person day-to-day associations; (3) physical activity interacted with sleep duration, with the protective effect of sleep on fatigue attenuated on more active days; and (4) in an exploratory reverse-direction analysis, evening energy (but not fatigue) showed a preliminary association with that night's nocturnal HRV that did not survive Bonferroni correction and requires replication.

Within-Person Sleep Effects and Between-Person HRV Associations

The distinction between within-person and between-person effects is conceptually critical for this study. The Mundlak models and rmcorr analyses converge on a consistent picture: day-to-day fluctuations in sleep duration predict day-to-day fluctuations in fatigue within the same individual. This is the most directly actionable finding: on nights when a person with long COVID sleeps longer than usual, they tend to feel less fatigued the next day.

The HRV-RMSSD finding is different in nature. Individuals with chronically higher HRV tend to report chronically better symptoms, but a given night's HRV-RMSSD does not predict whether that same individual will have a better or worse day of symptoms than their own average. This between-person difference may reflect stable individual characteristics, including underlying autonomic regulation capacity, severity of long COVID–related dysautonomia [34,35], baseline cardiovascular fitness, or other constitutional factors. This finding does not support HRV monitoring as a real-time, day-to-day symptom predictor in this population but does position HRV as a potentially useful biomarker of individual physiological reserve and overall symptom burden.

The pooled Spearman correlations (Multimedia Appendix 1) appeared to show much stronger HRV-symptom associations (ρ=−0.317 to −0.359 for fatigue, ρ=0.416 to 0.498 for energy). These were inflated by between-person variance: participants with higher average HRV also had lower average symptom burden, and, by treating 260 clustered observations as independent, this inflated both correlation coefficients and *P* values. The rmcorr and Mundlak analyses are methodologically appropriate for addressing the within-person scientific question and should be considered the primary analysis.

Exploratory Reverse-Direction Association

In an exploratory analysis of the reverse-direction pathway, evening energy but not fatigue showed a within-person association with that night's HRV-RMSSD (rmcorr *r*=0.327, *P*=.01; log-linear model β=0.0069, *P*=.02). This association was nominally significant but did not survive Bonferroni correction (α=.0050) and is therefore preliminary. If replicated, it would be consistent with the hypothesis that energy and fatigue reflect partially distinct processes: energy, assessed on a 0 to 100 scale, may capture perceived physiological vitality more sensitively than the 4-level fatigue scale, which was dominated by "moderate" responses (388/597, 65% of reports) and whose restricted variance may have limited its sensitivity (see *Limitations*). Energy levels on this scale cannot be equated with physiological capacity, and the association may partly reflect shared diurnal patterning, the 24-hour movement behavior framework [21], or other unmeasured factors. Replication in larger, adequately powered samples is required before any inference is drawn.

Physical Activity: Context-Dependent Effects

Neither activity level alone nor its direct association with symptoms was significant. However, the significant Sleep × Steps interaction (*P*=.02) suggests that the relationship between sleep and fatigue is context-dependent. On more active days, longer sleep conferred less symptomatic benefit than on less active days. This is consistent with PEM mechanisms [36], wherein greater exertion may place demands on physiological recovery that partially counteract the restorative function of sleep, and with behavioral evidence that perceived effort shapes activity decisions in individuals with fatigue [37]. These findings align with emerging evidence that sleep, sedentary time, and

XSL•FO
RenderX

physical activity should be considered as co-dependent daily behaviors influencing symptom outcomes [21]. The findings do not support direct exercise prescription, particularly given PEM concerns in this population. Low-burden, symptom-contingent movement approaches, including mind-body modalities such as Tai Chi, Qigong, or yoga, may offer more suitable future intervention directions [38-40], though the evidence specifically in long COVID requires dedicated investigation.

Activity guidance in long COVID may need to be personalized according to recent sleep quality and daily energy levels, in line with adaptive pacing principles [41].

Methodological Contributions

A key contribution of this study is the explicit use of symptom report timestamps to construct distinct prospective and reverse-direction analytical datasets, enabling temporally grounded bidirectional analysis. The use of rmcorr and Mundlak CLMMs represents an advance over standard pooled analyses and random-intercept-only models in repeated-measures ecological momentary assessment research.

Future studies in this space should preregister the distinction between within-person and between-person hypotheses, use rmcorr or mixed-model–based estimates as primary bivariate descriptors, and ideally include Mundlak specifications in all primary analyses.

Limitations

Several limitations should be considered when interpreting these findings. First, the small sample size (N=14) limits statistical power and the generalizability of the results. The cohort was predominantly female (n=12, 85.7%) and largely aged 45 to 64 years, reducing applicability to men and younger populations. Furthermore, the age range overlaps substantially with the perimenopausal and menopausal period, during which hormonal changes can independently influence sleep, autonomic function, and fatigue, representing a potential uncontrolled confounder. Participants were also recruited through support groups and patient and public involvement and engagement activities rather than clinician-confirmed diagnosis, which may have introduced heterogeneity in the sample. In addition, the absence of a healthy control group limits conclusions about whether the observed relationships are specific to long COVID–related fatigue.

Second, several methodological limitations should be acknowledged. HRV was derived from a consumer-grade Fitbit Inspire 3 using proprietary photoplethysmography algorithms, and spectral HRV measures, particularly LF power and the LF/HF ratio, have shown limited agreement with ECG-based measurements in previous validation studies. Consequently, these measures should be interpreted as indicative rather than definitive markers of autonomic function. The study also did not control for important factors known to influence HRV, including caffeine, alcohol, medications, or undiagnosed sleep disorders such as obstructive sleep apnea. Symptom reports were completed at participants' convenience rather than at fixed times, introducing potential recall bias and variability in the temporal alignment between symptoms, activity, and sleep. Additionally, excluding nights with poor HRV coverage (<50%) may have disproportionately removed periods of disturbed sleep, potentially introducing selection bias.

Finally, limitations related to outcome measurement and statistical modeling should be considered. Fatigue ratings were heavily clustered around moderate severity, reducing variability and potentially limiting the ability to detect day-to-day associations; future studies would benefit from validated multidimensional fatigue measures. Although the reverse-direction model showed acceptable improvement following log transformation, residuals remained imperfectly normal, indicating that model assumptions were only approximately satisfied.

Fatigue was modeled as a numeric outcome in the Sleep × Steps interaction model, whereas it is treated as ordinal in the primary cumulative link mixed models; because the observed fatigue scale has only three levels (mild, moderate, severe) with no assumed interval spacing, this continuous treatment is a simplification, and the interaction estimate should be interpreted with corresponding caution. In addition, because step counts were accumulated between sleep offset and a variable symptom report time, total steps are correlated with time awake at the point of reporting, and the Sleep × Steps interaction could therefore partly reflect a Sleep × Time Awake interaction.

Clinical Implications and Future Directions

These exploratory findings generate several hypotheses for future investigation. (1) Sleep duration and REM sleep are within-person day-to-day predictors of fatigue (the third association surviving correction, between sleep efficiency and next-day energy, is not carried forward here as it most plausibly reflects chance); sleep hygiene interventions targeting these metrics may offer symptom benefit. (2) HRV-RMSSD is a between-person correlate of average symptom burden, positioning it as a potential baseline biomarker rather than a real-time tracking tool. Whether autonomic-targeted interventions (HRV biofeedback, vagus nerve stimulation, slow breathing, mind-body practices) can improve between-person HRV levels and thereby reduce symptom burden in long COVID warrants experimental investigation [42,43]. (3) The preliminary, replication-dependent evening energy → nocturnal HRV association raises the hypothesis not yet supported by corrected evidence that evening relaxation strategies targeting subjective energy may support nocturnal autonomic recovery; testing this requires randomized experimental designs. (4) Activity pacing should incorporate recent sleep quality, with higher-exertion activities potentially better tolerated following nights of adequate sleep.

Randomized controlled trials manipulating sleep opportunity, HRV via biofeedback, or relaxation before sleep with crossover designs within the same individuals to leverage within-person variation would be the appropriate next step to test these hypotheses.

Conclusions

This exploratory longitudinal study identified bidirectional, temporally specific associations between sleep, nocturnal autonomic function, and daily symptoms in individuals with long COVID. Within-person analyses revealed that sleep

XSL•FO
RenderX

duration and REM sleep predict day-to-day fatigue fluctuations, while HRV-RMSSD differentiated individuals with better versus worse average symptom burden (a between-person rather than day-to-day predictor). A third association, between sleep efficiency and next-day energy, also survived correction but ran counter to expectation and most plausibly reflects chance. Evening energy, but not fatigue, showed a within-person association with subsequent nocturnal autonomic function. These findings underscore the importance of disaggregating within-person from between-person effects in ecological momentary assessment research and highlight the potential complementary roles of sleep tracking and HRV monitoring: the former for real-time daily prediction, the latter for characterizing physiological reserve in long COVID management. Experimental work is needed to clarify causal mechanisms.

Acknowledgments

The authors thank all participants for their time and engagement throughout the study. This work was supported by the National Institute for Health and Care Research Newcastle Biomedical Research Centre and is independent research. The authors are grateful to the Pausing-PCF team and patient and public involvement and engagement contributors for facilitating recruitment.

Funding

SDD was supported by the National Institute for Health and Care Research (NIHR) Newcastle Biomedical Research Centre based at The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle University and the Cumbria, Northumberland and Tyne and Wear (CNTW) NHS Foundation Trust. The research was also supported by NIHR Newcastle Clinical Research Facility (CRF) infrastructure funding. SDD was also supported by the Innovative Medicines Initiative 2 Joint Undertaking (IMI2 JU) project IDEA-FAST (grant 853981). SDD was supported by the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC) (grants EP/X031012/1 and EP/X036146/1). All opinions are those of the authors, not the funders.

Data Availability

Anonymized participant data and analysis code are available upon reasonable request from the corresponding author, subject to ethical approval. Participants did not consent to unrestricted data sharing; therefore, the data are available upon request rather than in a public repository.

Authors' Contributions

Conceptualization: NYA, MRB, KB, SDD
Data curation: NYA
Formal analysis: NYA
Funding acquisition: SDD
Investigation: NYA
Methodology: NYA
Project administration: NYA
Resources: NYA
Supervision: MRB, KB, SDD
Writing – original draft: NYA
Writing – review & editing: MRB, KB, SDD
All authors approved the final manuscript.

Conflicts of Interest

SDD is a shareholder in Enoda Health Ltd. This shareholding is unrelated to the work reported in this paper. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary tables S1-S3—standard cumulative link mixed model results (sensitivity analysis) and pooled Spearman correlations.
[DOCX File, 17 KB - cardio_v10i1e99630_app1.docx]

References

1. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine 2021 Aug;38:101019. [doi: 10.1016/j.eclinm.2021.101019] [Medline: 34308300]
2. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. Nat Med 2021 Apr;27(4):601-615. [doi: 10.1038/s41591-021-01283-z] [Medline: 33753937]

XSL•FO
RenderX

3. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet 2021 Jan 16;397(10270):220-232. [doi: 10.1016/S0140-6736(20)32656-8] [Medline: 33428867]
4. Mehandru S, Merad M. Pathological sequelae of long-haul COVID. Nat Immunol 2022 Feb;23(2):194-202. [doi: 10.1038/s41590-021-01104-y] [Medline: 35105985]
5. Gorna R, MacDermott N, Rayner C, et al. Long COVID guidelines need to reflect lived experience. Lancet 2021 Feb 6;397(10273):455-457. [doi: 10.1016/S0140-6736(20)32705-7] [Medline: 33357467]
6. Wong TL, Weitzer DJ. Long COVID and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)-a systematic review and comparison of clinical presentation and symptomatology. Medicina (Kaunas) 2021 Apr 26;57(5):418. [doi: 10.3390/medicina57050418] [Medline: 33925784]
7. Asarcikli LD, Hayiroglu M, Osken A, Keskin K, Kolak Z, Aksu T. Heart rate variability and cardiac autonomic functions in post-COVID period. J Interv Card Electrophysiol 2022 Apr;63(3):715-721. [doi: 10.1007/s10840-022-01138-8] [Medline: 35106678]
8. Barizien N, Le Guen M, Russel S, Touche P, Huang F, Vallée A. Clinical characterization of dysautonomia in long COVID-19 patients. Sci Rep 2021;11(1):14042. [doi: 10.1038/s41598-021-93546-5] [Medline: 34234251]
9. Hasty F, García G, Dávila H, Wittels SH, Hendricks S, Chong S. Heart rate variability as a possible predictive marker for acute inflammatory response in COVID-19 patients. Mil Med 2021 Jan 30;186(1-2):e34-e38. [doi: 10.1093/milmed/usaa405] [Medline: 33206183]
10. Mooren FC, Böckelmann I, Waranski M, et al. Autonomic dysregulation in long-term patients suffering from post-COVID-19 syndrome assessed by heart rate variability. Sci Rep 2023 Sep 22;13(1):15814. [doi: 10.1038/s41598-023-42615-y] [Medline: 37739977]
11. Chouchou F, Desseilles M. Heart rate variability: a tool to explore the sleeping brain? Front Neurosci 2014;8:402. [doi: 10.3389/fnins.2014.00402] [Medline: 25565936]
12. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996 Mar 1;93(5):1043-1065. [Medline: 8598068]
13. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Front Physio 2013;4(26). [doi: 10.3389/fphys.2013.00026] [Medline: 23431279]
14. Tobaldini E, Nobili L, Strada S, Casali KR, Braghiroli A, Montano N. Heart rate variability in normal and pathological sleep. Front Physiol 2013 Oct 16;4:294. [doi: 10.3389/fphys.2013.00294] [Medline: 24137133]
15. Aboagye NY, Germann M, Baker KF, Baker MR, Del Din S. Feasibility of predicting next-day fatigue levels using heart rate variability and activity-sleep metrics in people with post-COVID fatigue. Front Digit Health 2025;7:1689846. [doi: 10.3389/fdgth.2025.1689846] [Medline: 41341466]
16. Vivaldi G, Talaei M, Blaikley J, et al. Bidirectional relationship between sleep problems and long COVID: a longitudinal analysis of data from the COVIDENCE UK study. BMJ Open Resp Res 2025 Jun 15;12(1):e002506. [doi: 10.1136/bmjresp-2024-002506] [Medline: 40518292]
17. Board on the Health of Select Populations, Institute of Medicine, Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness: National Academies Press; 2015. [doi: 10.17226/19012] [Medline: 25695122]
18. Twomey R, DeMars J, Franklin K, Culos-Reed SN, Weatherald J, Wrightson JG. Chronic fatigue and postexertional malaise in people living with long COVID: an observational study. Phys Ther 2022 Apr 1;102(4):pzac005. [doi: 10.1093/ptj/pzac005] [Medline: 35079817]
19. Dehlia A, Guthridge MA. The persistence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) after SARS-CoV-2 infection: a systematic review and meta-analysis. J Infect 2024 Dec;89(6):106297. [doi: 10.1016/j.jinf.2024.106297] [Medline: 39353473]
20. Müller P, Achraf A, Zou L, Apfelbacher C, Erickson KI, Müller NG. COVID‐19, physical (in‐)activity, and dementia prevention. Alzheimers Dement (N Y) 2020;6(1):e12091. [doi: 10.1002/trc2.12091] [Medline: 33083514]
21. Jin X, Chen Y, Feng H, et al. Association of accelerometer-measured sleep duration and different intensities of physical activity with incident type 2 diabetes in a population-based cohort study. J Sport Health Sci 2024 Mar;13(2):222-232. [doi: 10.1016/j.jshs.2023.03.001] [Medline: 36871624]
22. Percutaneous auricular nerve stimulation for treating post-COVID fatigue (Pausing-PCF). COVID fatigue. URL: https://research.ncl.ac.uk/covidfatigue/ [accessed 2026-09-01]
23. Higuchi D, Takahashi Y, Tanaka S, Tomita Y. Validation of heart rate measured by the consumer-level wearable device Fitbit: a cross-sectional observational study. J Phys Ther Sci 2025 Nov;37(11):561-566. [doi: 10.1589/jpts.37.561] [Medline: 41209591]
24. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18. [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
25. Giurgiu M, von Haaren-Mack B, Fiedler J, et al. The wearable landscape: issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment. J Sport Health Sci 2025 Dec;14:101006. [doi: 10.1016/j.jshs.2024.101006] [Medline: 39491744]

XSL•FO
RenderX

26. Cao R, Azimi I, Sarhaddi F, et al. Accuracy assessment of Oura Ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: comprehensive analysis. J Med Internet Res 2022 Jan 18;24(1):e27487. [doi: 10.2196/27487] [Medline: 35040799]
27. Sarhaddi F, Kazemi K, Azimi I, et al. A comprehensive accuracy assessment of Samsung smartwatch heart rate and heart rate variability. PLoS One 2022;17(12):e0268361. [doi: 10.1371/journal.pone.0268361] [Medline: 36480505]
28. Heart rate time series. developer. URL: https://dev.fitbit.com/build/reference/web-api/heartrate-timeseries/ [accessed 2025-12-23]
29. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273. [doi: 10.2196/16273] [Medline: 31778122]
30. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;5:1040. [doi: 10.3389/fpsyg.2014.01040] [Medline: 25324790]
31. FatigueSense. URL: https://www.fatiguesense.com/ [accessed 2026-08-29]
32. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017;8:456. [doi: 10.3389/fpsyg.2017.00456] [Medline: 28439244]
33. Mundlak Y. On the pooling of time series and cross section data. Econometrica 1978;46(1):69-85. [doi: 10.2307/1913646]
34. Dani M, Dirksen A, Taraborrelli P, et al. Autonomic dysfunction in 'long COVID': rationale, physiology and management strategies. Clin Med (Lond) 2021 Jan;21(1):e63-e67. [doi: 10.7861/clinmed.2020-0896] [Medline: 33243837]
35. Tabacof L, Wood J, Breyman E, et al. Dysautonomia, but not cardiac dysfunction, is common in a cohort of individuals with long COVID. J Pers Med 2023 Nov 14;13(11):1606. [doi: 10.3390/jpm13111606] [Medline: 38003921]
36. Vernon SD, Hartle M, Sullivan K, et al. Post-exertional malaise among people with long COVID compared to myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Work 2023;74(4):1179-1186. [doi: 10.3233/WOR-220581] [Medline: 36911963]
37. Maltagliati S, Fessler L, Yu Q, et al. Effort minimization: a permanent, dynamic, and surmountable influence on physical activity. J Sport Health Sci 2025 Dec;14:100971. [doi: 10.1016/j.jshs.2024.100971] [Medline: 39233203]
38. Zou L, Sasaki JE, Wei GX, et al. Effects of mind-body exercises (Tai Chi/yoga) on heart rate variability parameters and perceived stress: a systematic review with meta-analysis of randomized controlled trials. J Clin Med 2018 Oct 31;7(11):404. [doi: 10.3390/jcm7110404] [Medline: 30384420]
39. Zou L, SasaKi JE, Wang H, Xiao Z, Fang Q, Zhang M. A systematic review and meta-analysis of Baduanjin qigong for health benefits: randomized controlled trials. Evid Based Complement Alternat Med 2017;2017:4548706. [doi: 10.1155/2017/4548706] [Medline: 28367223]
40. Liu J, Yang Y, Li C, et al. Effects of mind-body qigong exercise on overall health, fatigue/sleep, and cognition in older Chinese immigrants in the US: an intervention study with control. J Aging Res 2024;2024:2481518. [doi: 10.1155/2024/2481518] [Medline: 38333772]
41. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet 2011 Mar 5;377(9768):823-836. [doi: 10.1016/S0140-6736(11)60096-2] [Medline: 21334061]
42. Lehrer P, Kaur K, Sharma A, et al. Heart rate variability biofeedback improves emotional and physical health and performance: a systematic review and meta analysis. Appl Psychophysiol Biofeedback 2020 Sep;45(3):109-129. [doi: 10.1007/s10484-020-09466-z] [Medline: 32385728]
43. Bonaz B, Sinniger V, Pellissier S. Vagal tone: effects on sensitivity, motility, and inflammation. Neurogastroenterol Motil 2016 Apr;28(4):455-462. [doi: 10.1111/nmo.12817] [Medline: 27010234]

Abbreviations

CLMM: cumulative link mixed model
ECG: electrocardiogram
HF: high-frequency
HRV: heart rate variability
ICC: intraclass correlation coefficient
LF: low-frequency
PEM: postexertional malaise
REM: rapid eye movement
rmcorr: repeated-measures correlation
RMSSD: root mean square of successive differences
VIF: variance inflation factor

XSL•FO
RenderX

Edited by A Coristine; submitted 27.Apr.2026; peer-reviewed by B Zhai, L Zou; revised version received 14.Aug.2026; accepted 18.Aug.2026; published 14.Sep.2026.

Please cite as:
Aboagye NY, R Baker M, Baker K, Del Din S
Bidirectional Associations Between Autonomic Function, Sleep Quality, and Daily Fatigue and Energy Symptoms in Long COVID: Longitudinal Digital Health Study
JMIR Cardio 2026;10:e99630
URL: https://cardio.jmir.org/2026/1/e99630
doi:10.2196/99630

© Nana Yaw Aboagye, Mark R Baker, Kenneth Baker, Silvia Del Din. Originally published in JMIR Cardio (https://cardio.jmir.org), 14.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bidirectional Automated Texting for Cardiovascular Health Among People Living With HIV: Observational Cohort Analysis of a Stepped-Wedge Cluster Randomized Controlled Trial

Mechelle Sanders[1], PhD; Donald Harrington[2], MBA, MS; Emma Sass[1], MPH; Marie Thomas[1]; Tameir Holder[3], MPH; Yiqi Tian[3], MS; Brent Johnson[4], PhD; Andrea Cassells[3], MPH; Jonathan N Tobin[3,5], PhD; Kevin Fiscella[1], MD, MPH

[1]Department of Family Medicine, University of Rochester Medical Center, Rochester, NY, United States

[2]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, United States

[3]Clinical Directors Network, New York, NY, United States

[4]Department of Biostatistics, University of Rochester Medical Center, Rochester, NY, United States

[5]The Rockefeller University, The Rockefeller University Center for Clinical and Translational Science, New York, NY, United States

Corresponding Author:
Mechelle Sanders, PhD
Department of Family Medicine, University of Rochester Medical Center, Rochester, NY, United States

Abstract

Background: Feasible and potentially scalable strategies are needed to address the growing cardiovascular disease (CVD) risk among people living with HIV. Bidirectional automated texting (BAT) programs that remind and encourage adherence to evidence-based CVD-reducing interventions represent a potentially scalable strategy, but data on their feasibility are lacking.

Objective: The goal of the study was to determine whether participant sociodemographic factors and technological constraints influenced engagement with a BAT CVD prevention program by people living with HIV. We conducted an observational cohort analysis embedded within a stepped-wedge cluster randomized trial.

Methods: The BAT program was designed to address the "Million Hearts" ABCS (aspirin therapy, blood pressure control, cholesterol management, and smoking cessation) of cardiovascular health. The parent study was a stepped-wedge randomized trial that rolled out in 3 wedges across 8 practice sites that provided care to people living with HIV. Participants received and could engage with the text messages weekly using their own phones during the study period. We used a zero-inflated negative binomial model to identify factors associated with participants sending text messages during the study.

Results: Of the 471 participants, 94% owned a smartphone capable of text messaging, and 70% reported monthly incomes less than US $1500. Overall, 60.3% (n=284) engaged with the BAT program at least once. Regarding texting behavior, participants aged ≥65 years were more likely to send a text than those aged <50 years (P=.047), although age did not influence the number of texts sent. White participants showed lower texting intensity than Black participants (incidence rate ratio 0.69; P=.04).

Conclusions: Overall, 60.3% (n=284) of the participants in the study engaged with BAT at least once. The BAT intervention for ABCS appears to be a feasible intervention for people living with HIV. Only a few factors were associated with sending a text or with the number of text messages sent.

Trial Registration: ClinicalTrials.gov NCT03902431; https://clinicaltrials.gov/study/NCT03902431

(JMIR Cardio 2026;10:e69098) doi:10.2196/69098

KEYWORDS

mobile health; mHealth; digital intervention; text messaging; HIV; cardiovascular disease

Introduction

Use of highly effective antiretroviral therapy has dramatically reduced deaths from AIDS-related causes, yielding an aging population among people living with HIV [1,2]. Cardiovascular disease (CVD) risk is up to 2 times greater among people living with HIV compared with the general population [3,4]. Bidirectional automated texting (BAT) has been successfully used for multiple types of health interventions, including tobacco cessation, blood pressure control, cancer screening, medication adherence, and HIV treatment [5-10]. BAT allows users to respond to confirm medical appointments, answer questions, report symptoms, and share health updates in real time or at their convenience, enabling efficient and ongoing asynchronous communication between patients and their health care team [5-7]. BAT can also be used to collect health information through text messages. It is a potentially scalable tool for communicating with patients outside the clinical encounter

XSL•FO
RenderX

while enabling the provision of tailored information based on responses to prior text messages [8].

A systematic review found that text-messaging programs for CVD can help people better manage their heart disease risk factors [9]. Given this, the growing risk of CVD among people living with HIV may be substantively reduced by using evidence-based BAT programs that remind and encourage adherence to evidence-based CVD-reducing interventions [10]. Compared with other types of engagement, texting offers potentially higher reach relative to cost and is scalable due to its low marginal costs once established [10-13]. Beyond the general efficacy, specific trials have also explored the practical utility and user acceptance of texting interventions for preventing and managing CVD [14,15]. For example, in one study, participants received 1 text per week about general CVD knowledge, blood pressure control, medication adherence, physical activity, healthy diet, and smoking cessation (only for smokers). The findings highlighted strong user satisfaction, with nearly all participants reporting that the text messages were useful (96.1%), easy to understand (98.8%), and appropriate in frequency (93.8%). Participants also reported a high willingness to receive future messages (94.8%) [16].

Several studies have shown that text messaging can improve medication adherence and appointment attendance [17-21]. However, evidence for its impact on sexually transmitted infection and HIV prevention and treatment remains mixed [22,23]. McLaughlin et al [24] proposed a framework for creating educational text messages around primary prevention of CVD in people living with HIV, but randomized trials are still needed to confirm its effectiveness. To address this gap, we analyzed feasibility data from an observational cohort nested within a stepped-wedge trial to identify factors influencing engagement among people living with HIV with CVD prevention text messages.

Methods

Population

The HIV ABCS (aspirin therapy, blood pressure control, cholesterol management, and smoking cessation) trial, which was part of PRECluDE (Implementation Research to Develop Interventions for People Living With HIV), a consortium of National Heart, Lung, and Blood Institute funded trials, has been previously described (ClinicalTrials.gov NCT03902431) [25,26]. Participants were recruited from 8 geographically diverse sites across 3 locations: 4 in New York City, New York; 2 in Rochester, New York; and 2 in Dallas, Texas. These sites included academic medical centers, freestanding HIV practices, and Federally Qualified Health Centers (FQHCs). Participants were eligible if they had English or Spanish language proficiency sufficient to read text messages and a personal mobile phone. We enrolled 471 participants aged 40 to 75 years (by July 1, 2018) with a 10-year atherosclerotic CVD (ASCVD) risk of at least 5% without known cardiovascular or cerebrovascular disease (eg, stroke, myocardial infarction, and coronary artery disease), based on eligibility criteria from the ASCVD risk calculator [27].

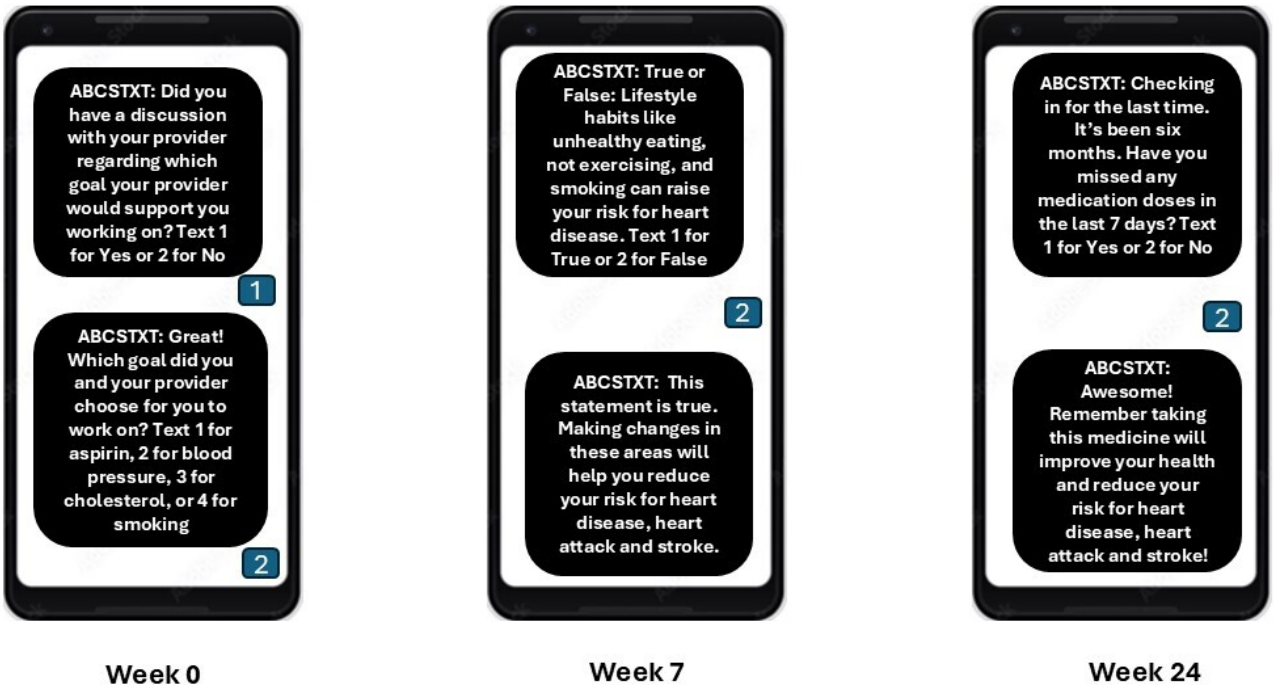
Intervention

This study was an observational cohort nested within the parent stepped-wedge trial. The parent stepped-wedge trial was conducted between 2019 and 2023. This design used a multicomponent strategy to promote the uptake of the US Department of Health and Human Services' "Million Hearts" ABCS campaign for cardiovascular health [28]. This design involved 8 sites divided into 3 wedges, with each wedge separated by 6 to 8 months.

For the patient-facing component, BAT messages were delivered by a commercial vendor (Mosio). To develop the BAT for CVD prevention for people living with HIV, we generated a library of evidence-based text messages for each of the ABCS. The messages included potential behavior change techniques, knowledge about the ABCS, personal affirmation, self-efficacy, adherence support, and the importance of goal commitment (Figure 1). Mosio sent messages in either English or Spanish, depending on participants' language preferences.

XSL•FO
RenderX

Figure 1. Examples of goal-setting text messages sent to the participants in the HIV ABCS (aspirin therapy, blood pressure control, cholesterol management, and smoking cessation) and their responses at different time points.



Participants engaged with these text messages for a maximum of 6 months, regardless of their assigned wedge. Message frequency was initially set at 1 to 3 texts per day for the first 4 months, then tapered to 1 to 2 texts per week. This frequency was also adaptive, with additional texts sent based on participant responses to prompts, such as indicating a need for more information or incorrect answers to CVD knowledge checks. This was embedded in the program's branching logic. To encourage sustained engagement throughout the 6-month texting period, participants were eligible for 3 US $5 incentives (up to US $15 overall) if they responded to specific text messages that were randomly selected from the automated message library [29].

Demographic and Psychosocial Variables

Research assistants collected data from participants on sex at birth, race, ethnicity, preferred language, and income. Participants also reported their chronic conditions (hypertension or diabetes) and completed validated scales: the Patient Activation Measure (PAM), the Confusion, Hubbub, and Order Scale (CHAOS), and the Subjective Numeracy Scale (SNS) [30-32]. The PAM assessed patients' knowledge, skills, and confidence in managing their health. The CHAOS scale evaluated household environment disorganization and its potential impact on texting behaviors, while the SNS measured patients' abilities to understand and use health-related information effectively. Together, these instruments provided a comprehensive assessment of factors influencing self-management and texting engagement. Both the PAM and SNS have been evaluated in similar populations [33,34].

Smartphone and Text Message Use and Skills

We developed 8 questions to evaluate participants' smartphone use and skills, covering ownership, length of use, acquisition type, self-rated skill level, frequency of service interruption, ability to text during interruptions, and time since the last interruption. For instance, participants reported years of smartphone ownership or use, with options ranging from "0 - 3 years" to "more than 7 years." Participants also rated their smartphone skill from "Beginner" (limited to calls and texts) to "Expert" (full proficiency, including settings). Phone acquisition options included "Purchase," "Gift," or "Government Plan." Participants indicated how they paid for their data service plan (eg, postpaid, prepaid, limited minutes, or other options). The frequency of phone or data service interruptions was assessed with options ranging from "every month by the first week" to "never." Finally, the survey asked about unlimited texting capability and the ability to receive texts during service interruptions, with "Yes," "No," and "Unsure" as response options.

Texting Engagement

Previous studies of text-messaging interventions have shown that outcomes often follow a dose-response pattern, where greater engagement can lead to larger improvements in the variables of interest [35-37]. To assess patient engagement with BAT, we counted the number of text messages patients sent. We created two engagement measures: (1) a binary indicator of any engagement (sent at least 1 message vs no messages) and (2) a continuous count of messages among participants who sent at least 1 message.

ABCS Goal Choice

We identified the participants' ABCS goal based on their report within the BAT system. Participants specifically reported the goal that they had discussed and mutually agreed upon with their clinician.

Data Analysis

We performed all analyses using R (version 4.4.2; R Foundation for Statistical Computing). All statistical tests were 2-tailed and

XSL•FO
RenderX

based on a 5% significance level. We calculated descriptive statistics and presented these as means and SDs where appropriate.

We used a zero-inflated negative binomial (ZINB) model to assess factors associated with participants sending a text. Our model accounted for excess zeros (eg, never sending a text) and overdispersion. Covariates for the model were selected using a 2-step process: first, variables were screened for predictive importance using Bayesian additive regression trees (BART) with the threshold of 0.4; second, the final model included variables identified by BART as well as standard sociodemographic adjusters (eg, age, race, and sex) important to the study design and supported by existing literature. Ethnicity was screened out by the BART procedure and was not retained in the final model owing to its collinearity with race. The count model examined the frequency of texts sent among those who sent at least 1 text (incidence rate ratio), while the inflation model predicted the likelihood of never sending a text versus sending any texts (odds ratio).

Ethical Considerations

The study was approved by the institutional review board at the University of Rochester Medical Center (approval number STUDY00003679), with reliance review at each of the participating sites. Informed consent was obtained from each participant. Participants could receive up to US $80 for this study: US $40 for the baseline visit, US $10 for the second visit, US $15 for the third visit, and up to an additional US $15 (3 US $5 cards) for responding to randomly incentivized text messages, which were provided at the second or third visit.

Results

Participants

The participants were middle-aged, that is, 55 years or older (336/470, 71.5%), male (321/471, 68.2%), and Black (273/470, 58.1%). Among participants with available income data, 48.7% (211/433) reported a monthly household income of less than US $1000 per month (Table 1).

XSL•FO
RenderX

Table . Baseline participant characteristics[a].

Characteristics	Participants
Demographics	
Age (y; n=470), n (%)	
40 - 49	50 (10.6)
50 - 54	84 (17.9)
55 - 59	134 (28.5)
60 - 64	114 (24.3)
≥65	88 (18.7)
Education (n=443), n (%)	
Less than high school	107 (24.2)
High school graduate	141 (31.8)
Higher than high school	195 (44)
Monthly household income (US $; n=433), n (%)	
<1000	211 (48.7)
1000 - 1499	89 (20.6)
1500 - 1999	55 (12.7)
>2000	78 (18)
Sex at birth (n=471), n (%)	
Male	321 (68.2)
Female	149 (31.6)
Decline to state	1 (0.2)
Race (n=470), n (%)	
Black or African American	273 (58.1)
White	136 (28.9)
Other	61 (13)
Ethnicity (Hispanic or Latino; n=434), n (%)	87 (20)
Clinical characteristics	
Blood pressure at baseline (mm Hg), mean (SD)	
Systolic (n=470)	135.32 (18.2)
Diastolic (n=464)	82.52 (10.7)
Treated for hypertension (n=470), n (%)	293 (62.3)
Has diabetes (n=470), n (%)	154 (32.8)
Smoking status (n=470), n (%)	
Current	195 (41.5)
Never	147 (31.3)
Former	128 (27.2)
Regular aspirin use (n=465), n (%)	104 (22.4)
Treated with statin (n=466), n (%)	266 (57.1)
Total cholesterol (mg/dL; n=470), mean (SD)	186.2 (37.5)
Survey measures, mean (SD)	
Health literacy score (n=412)	4.48 (1.3)
Subjective Numeracy Scale score (1-6; n=446)	3.94 (1.0)
Confusion, Hubbub, and Order Scale score (0 - 30; n=433)	13.98 (4.3)

XSL•FO
RenderX

Characteristics	Participants
Patient Activation Measure raw score (13-52; n=447)	43.55 (5.9)
Technology access and skills	
Own or have access to a smartphone (n=446), n (%)	417 (93.5)
Duration of owning or using a smartphone (y; n=439), n (%)	
0 - 3	141 (32.1)
4 - 7	103 (23.5)
>7	195 (44.4)
Smartphone acquisition (n=446), n (%)	
Purchased	359 (80.5)
Gift	72 (16.1)
Government plan	15 (3.4)
Self-rated smartphone skill (n=439), n (%)	
Beginner (calls and text only)	117 (26.7)
Advanced beginner (calls and text and internet search)	60 (13.7)
Intermediate (calls and texts, internet search, and downloading apps)	110 (25.1)
Advanced (calls and texts, internet search, downloading apps, and changing settings and passwords)	75 (17.1)
Expert (all of the skills and more)	77 (17.6)
Phone or data service interruption frequency (n=444), n (%)	
Every month, first week	34 (7.8)
Every month, second week	8 (1.8)
Every month, third week	11 (2.5)
Rarely	87 (19.6)
Never	304 (68.5)
Unlimited texting capability (n=446), n (%)	383 (85.9)
Able to receive texts when service interrupted (n=441), n (%)	211 (47.8)
Service interrupted in past 4 months (n=314), n (%)	31 (9.9)
ABCS[b] goal choice (n=448), n (%)	
Aspirin	28 (6.3)
Blood pressure	136 (30.4)
Cholesterol	169 (37.8)
Smoking	115 (25.8)
Primary outcome measure	
Did the participant send a text message? (n=471), n (%)	284 (60.3)
Message count (among those that sent a text; n=284), mean (SD)	29.26 (34.2)
Participants per site (n=471), n (%)	
Site 1	109 (23.1)
Site 2	43 (9.1)
Site 3	57 (12.1)
Site 4	40 (8.5)
Site 5	31 (6.7)
Site 6	42 (8.9)
Site 7	110 (23.4)

XSL•FO
RenderX

Characteristics	Participants
Site 8	39 (8.3)

[a]The sample sizes vary across variables because not all participants provided responses for every measure, including missing data in electronic health records.

[b]ABCS: aspirin therapy, blood pressure control, cholesterol management, and smoking cessation.

A total of 93.5% (417/446) of participants owned smartphones and 32.1% (141/439) had owned or used one for ≤3 years. Most participants (359/446, 80.5%) purchased their phone, whereas 3.4% (15/446) received it through a government plan. Most participants (304/444, 68.5%) never had service interruptions. Self-rated smartphone skills varied, with 26.7% (117/439) reporting being beginners and 34.6% (152/439) as advanced or expert (Table 1).

ZINB Model

Out of a total of 39 potential variables, BART selected 5 (12.8%) variables as important predictors of texting: smoking status, clinical practice site, self-rated smartphone skill, length of smartphone use, and final goal (ABCS), but none of the 4 survey measures. Together with age, sex, and race, the final model included 8 independent predictors. We fit the ZINB model to the text count outcome and the above 8 key covariates, after excluding 38 participants due to missing data (Table 2).

XSL•FO
RenderX

Table . Zero-Inflation model: predictors of texting engagement[a].

Predictor	Never texted, OR[b] (95% CI)	*P* value	Text count, IRR[c] (95% CI)	*P* value
Age (y)				
<50	Reference	—[d]	—	—
50 - 54	0.63 (0.24-1.67)	.36	1.13 (0.61-2.09)	.71
55 - 59	0.56 (0.23-1.40)	.21	1.28 (0.72-2.27)	.44
60 - 64	0.78 (0.30-2.00)	.61	0.87 (0.48-1.56)	.64
≥65	0.33 (0.11-0.99)	.047	0.80 (0.44-1.47)	.48
Sex				
Female	Reference	—	—	—
Male	0.78 (0.42-1.46)	.44	0.81 (0.58-1.12)	.22
Race				
Black or African American	Reference	—	—	—
White	0.65 (0.30-1.44)	.29	0.69 (0.49-0.98)	.04
Other	1.14 (0.51-2.54)	.74	0.84 (0.54-1.31)	.44
Smoking status				
Current	Reference	—	—	—
Former	1.33 (0.59-2.99)	.49	1.18 (0.77-1.79)	.45
Never	1.32 (0.61-2.82)	.48	1.02 (0.67-1.51)	.92
Study site				
Site 1	Reference	—	—	—
Site 2	0.91 (0.38-2.19)	.84	0.55 (0.28-1.08)	.08
Site 3	0.22 (0.08-0.59)	.003	0.34 (0.18-0.62)	<.001
Site 4	0.14 (0.04-0.46)	.001	0.36 (0.20-0.63)	<.001
Site 5	0.54 (0.20-1.48)	.23	0.67 (0.32-1.39)	.28
Site 6	0.20 (0.07-0.59)	.004	0.76 (0.42-1.37)	.36
Site 7	0.08 (0.03-0.21)	<.001	1.23 (0.81-1.89)	.33
Site 8	0.09 (0.02-0.40)	.002	0.59 (0.34-1.02)	.06
Self-rated smartphone skill				
Beginner	Reference	—	—	—
Advanced beginner	0.76 (0.33-1.79)	.53	1.19 (0.76-1.84)	.45
Intermediate	0.70 (0.33-1.47)	.35	1.33 (0.88-2.00)	.18
Advanced	0.74 (0.30-1.83)	.52	1.19 (0.75-1.89)	.46
Expert	0.39 (0.14-1.09)	.07	1.20 (0.77-1.85)	.42
Length of smartphone use (y)				
<4	Reference	—	—	—
4 - 7	0.91 (0.45-1.83)	.79	1.27 (0.87-1.86)	.22
>7	0.72 (0.36-1.43)	.34	1.04 (0.74-1.46)	.81
ABCS[e] goal				
Cholesterol	Reference	—	—	—
Aspirin	0.25 (0.07-0.83)	.02	1.45 (0.83-2.51)	.19
Blood pressure	0.86 (0.46-1.61)	.63	0.90 (0.65-1.25)	.53

XSL•FO
RenderX

Predictor	Never texted, OR[b] (95% CI)	*P* value	Text count, IRR[c] (95% CI)	*P* value
Smoking	1.39 (0.59-3.25)	.45	1.06 (0.68-1.64)	.80

[a]The point estimate for the natural logarithm of the size parameter in the statistical model was −0.03 (95% CI −0.26 to 0.19; *P*=.77).

[b]OR: odds ratio.

[c]IRR: incidence rate ratio.

[d]Reference category (zero inflation).

[e]ABCS: aspirin therapy, blood pressure control, cholesterol management, and smoking cessation.

The odds of ever texting were 3-fold higher for participants aged ≥65 years than those aged <50 years (*P*=.047). However, age was not associated with the number of texts sent among participants who sent at least 1 text. Regarding race, we observed no statistically significant impact on whether a participant ever sent a text. However, among those who texted, White participants sent approximately 31% fewer text messages than Black participants (*P*=.04). Participants at 5 sites were more likely to send at least 1 text compared with those at the reference site. However, among participants who texted at least once, those from two of these same sites sent fewer text messages.

Discussion

Principal Results

This study explored factors affecting people living with HIV engagement with a BAT CVD prevention texting program among people living with HIV. In our ZINB model, older age, HIV care site, and selecting aspirin as a goal were associated with texting participation. The association between site and texting engagement warrants further investigation. It is not clear why, after controlling for patient-level factors, participants at 5 sites were more likely to send at least 1 text, and why those at 2 of these sites who texted at least once sent *fewer* text messages. This unexpected finding may serve as a surrogate for unmeasured site-specific factors such as differences in staffing, changes in on-site technical support for texting, prior digital health experience, or the quality of study implementation. These unmeasured factors may have been stronger predictors of engagement than patient-level demographic characteristics. The association with aspirin may reflect the complexity of the 2016 US Preventive Services Task Force's aspirin recommendations: "The USPSTF recommends initiating low-dose aspirin use for the primary prevention of cardiovascular disease (CVD) and colorectal cancer in adults aged 50 to 59 years who have a 10% or greater 10-year CVD risk, are not at increased risk for bleeding, have a life expectancy of at least 10 years, and are willing to take low-dose aspirin daily for at least 10 years [38]." Participants selecting aspirin may have texted more due to this added complexity. While our data cannot definitively assess these explanations, future research incorporating qualitative methods could investigate these contextual factors more deeply.

Comparison With Prior Work

Our results align with previous studies demonstrating BAT is feasible and readily used among historically lower digitally literate participants, that is, those who are older, have low income, and have less education [39]. Our results also reflect the marked increase in smartphone ownership across all demographic groups, consistent with broader trends reported in the literature [40]. Our analysis included primarily participants who did not report monthly interruptions in mobile phone service. This approach allowed us to examine patient engagement within a cohort largely free from connectivity challenges, which previous studies have identified as a significant barrier [41,42]. Building on prior work by our group and others that indicates the feasibility of providing digital literacy training to low-income people living with HIV [43,44], our findings further support the potential for technology-assisted interventions to increase engagement with care. This suggests a promising avenue for future intervention development, potentially increasing engagement with care, and facilitating the evaluation of both effectiveness and implementation outcomes.

While race did not significantly influence the likelihood of a participant sending any text, we observed that among those who did text, White participants sent fewer text messages than Black participants. This finding is consistent with earlier data from 2010 showing that Black and Hispanic individuals who use text messaging text more on average than White individuals, with a median of 10 text messages per day for Black and Hispanic individuals compared with 5 text messages per day for White individuals [10]. To mitigate language barriers a priori, our research staff who enrolled Spanish-speaking patients were bilingual, and BAT messages were translated and back-translated into Spanish. However, our subsample of Spanish-speaking patients was modest, and we cannot rule out the need for further cultural or linguistic adaptations to the text messages.

Limitations and Strengths

One limitation of this study is the lack of qualitative data at the patient, clinician, and site levels. We were limited to quantitative information and may have missed important contextual factors that could help explain the differences observed across sites. For example, at the start of the intervention, participants were able to self-select the best time of day to receive texts. However, participants' preferred texting times may have changed over time (eg, due to changes in employment hours), potentially limiting their ability to engage with the intervention. Additionally, it is possible that participants saw the texts, read them, and decided not to engage with the text. Lack of engagement does not necessarily preclude them from following through with their ABCS goals. In univariate analyses, we observed that participants experiencing service interruptions were less likely to send messages (*P*<.001). However, neither service interruption variable was important in the BART variable selection procedure after adjusting for the other key variables.

XSL•FO
RenderX

Importantly, although we adopted a complete-case analysis in the ZINB model, the BART procedure performs feature selection without this assumption under a missing at random assumption and, therefore, did not remove records with missing values when selecting important features. However, the BART method for handling missing data is biased toward the null, which implies that BART would tend to shrink a nonnegligible correlation closer to zero in the presence of missing data [45]. Consequently, our measure of engagement may underestimate true engagement among those facing connectivity challenges. The prevalence of service interruptions itself may be underestimated if those most affected were also less likely to report them. The study used modest random monetary incentives to increase texting engagement, with participants able to receive an additional US $15 for responding to randomly incentivized messages during the 6-month intervention. This approach may have influenced their motivation to text, and it is unclear whether this motivation would persist in a real-world setting without such incentives. Moreover, we were unable to control for these random incentives in our analysis, making them a potential confounder for the number of texts sent. A separate limitation is that we were unable to control for the adaptive logic of the text messaging (eg, additional prompts based on responses), which may have biased the estimated associations with texting frequency or engagement.

Despite these limitations, the strengths of our study include the availability of both Spanish and English versions of the text messages, a geographically diverse participant population, multiple types of care settings (eg, FQHCs and academic health centers), and a large sample size that enhances the generalizability of the findings.

Conclusions

A BAT intervention for the ABCS of heart health appears to be a promising tool for engaging people living with HIV. Overall, 60.3% (284/471) of the participants in the study were able to engage with the texting platform at least once. Surprisingly, few factors were associated with sending a text or with the number of text messages sent.

BAT is feasible to implement among low-income people living with HIV. Future studies should examine contextual factors that may have influenced engagement but were not captured in our research, particularly the role of personal assistance. Given the increasing reliance on text messaging by health systems for patient communication (eg, scheduling, reminders, and results), future efforts should explore how BAT interventions can be integrated into clinical workflows to reinforce CVD prevention, particularly among patients with low incomes.

Acknowledgments

The authors gratefully acknowledge the invaluable contributions of their collaborating partners and organizations whose participation made this investigation possible: Joseph P. Addabbo Health Center (Queens, New York); Morris Heights Health Center (Bronx, New York); Brownsville Multi-specialty Practice (Brooklyn, New York); Family Health Centers at New York University Langone–Sunset Terrace Health Center (Brooklyn, New York); Bluitt-Flowers Health Clinic (Dallas, Texas); Amelia Court Clinic (Dallas, Texas); University of Rochester Medical Center AIDS Center (Rochester, New York); and Trillium Health (Rochester, New York). The authors also extend their appreciation to the National Heart, Lung, and Blood Institute Center for Translation Research and Implementation Science staff, including Cheryl Boyce, PhD; Mary Masterson, PhD; Rebecca Roper, MS, MPH, EdD; and Veronica Tutse-Tonwe, MSHS, PMP, for their guidance and support. Finally, the authors thank the members of the Data and Safety Monitoring Board—Anita Vaughn, MD; Katherine Freeman, DrPH; and David Young—for their oversight and commitment throughout the study. Grammarly was used for copyediting assistance with the final manuscript, and Qualified Health AI was used to confirm that the authors' responses to the reviewers' comments were adequate.

Funding

This study was funded by the National Heart, Lung, and Blood Institute through grant U01HL142107, beginning in 2018. The project described in this publication was also supported by the University of Rochester Clinical and Translational Science Award (CTSA), award UL1 TR002001, from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant confidentiality but can be made available by the corresponding author upon reasonable request.

Authors' Contributions

KF and JNT contributed to funding acquisition. MS, KF, and JNT contributed to the conceptualization and methodology of this study. BJ and DH contributed to data curation and investigation. MS contributed to writing the original draft. MS, KF, JNT, ES, TH, MT, YT, and AC contributed to analysis review and validation. AC, ES, TH, and MT contributed to project administration. All authors contributed to writing, review, and editing and approved the submitted manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Trickey A, Sabin CA, Burkholder G, et al. Life expectancy after 2015 of adults with HIV on long-term antiretroviral therapy in Europe and North America: a collaborative analysis of cohort studies. Lancet HIV 2023 May;10(5):e295-e307. [doi: 10.1016/S2352-3018(23)00028-0] [Medline: 36958365]
2. Trickey A, McGinnis K, Gill MJ, et al. Longitudinal trends in causes of death among adults with HIV on antiretroviral therapy in Europe and North America from 1996 to 2020: a collaboration of cohort studies. Lancet HIV 2024 Mar;11(3):e176-e185. [doi: 10.1016/S2352-3018(23)00272-2] [Medline: 38280393]
3. Alonso A, Barnes AE, Guest JL, Shah A, Shao IY, Marconi V. HIV infection and incidence of cardiovascular diseases: an analysis of a large healthcare database. J Am Heart Assoc 2019 Jul 16;8(14):e012241. [doi: 10.1161/JAHA.119.012241] [Medline: 31266386]
4. Losina E, Hyle EP, Borre ED, et al. Projecting 10-year, 20-year, and lifetime risks of cardiovascular disease in persons living with human immunodeficiency virus in the United States. Clin Infect Dis 2017 Oct 15;65(8):1266-1271. [doi: 10.1093/cid/cix547] [Medline: 28605504]
5. Ingersoll K, Dillingham R, Reynolds G, et al. Development of a personalized bidirectional text messaging tool for HIV adherence assessment and intervention among substance abusers. J Subst Abuse Treat 2014 Jan;46(1):66-73. [doi: 10.1016/j.jsat.2013.08.002] [Medline: 24029625]
6. Donovan G, Hall N, Ling J, Smith F, Wilkes S. Influencing medication taking behaviors using automated two‐way digital communication: a narrative synthesis systematic review informed by the Behavior Change Wheel. Br J Health Psychol 2022 Sep;27(3):861-890. [doi: 10.1111/bjhp.12580]
7. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. J Med Internet Res 2016 Apr 19;18(4):e86. [doi: 10.2196/jmir.5425] [Medline: 27095386]
8. Abroms LC, Whittaker R, Free C, Mendel Van Alstyne J, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107. [doi: 10.2196/mhealth.4917] [Medline: 26690917]
9. Calderon Martinez E, Zachariah Saji S, Salazar Ore JV, et al. Text messages as a tool to improve cardiovascular disease risk factors control: a systematic review and meta-analysis of randomized clinical trials. BMC Public Health 2025 Apr 4;25(1):1284. [doi: 10.1186/s12889-025-21818-0] [Medline: 40186180]
10. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
11. Leong KC, Chen WS, Leong KW, et al. The use of text messaging to improve attendance in primary care: a randomized controlled trial. Fam Pract 2006 Dec;23(6):699-705. [doi: 10.1093/fampra/cml044] [Medline: 16916871]
12. Chen ZW, Fang LZ, Chen LY, Dai HL. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: a randomized controlled trial. J Zhejiang Univ Sci B 2008 Jan;9(1):34-38. [doi: 10.1631/jzus.B071464] [Medline: 18196610]
13. Siopis G, Moschonis G, Eweka E, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with hypertension: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e144-e159. [doi: 10.1016/S2589-7500(23)00002-X] [Medline: 36828607]
14. Shariful Islam SM, Farmer AJ, Bobrow K, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019;6(2):e001017. [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
15. Ricci-Cabello I, Bobrow K, Islam SM, et al. Examining development processes for text messaging interventions to prevent cardiovascular disease: systematic literature review. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e12191. [doi: 10.2196/12191] [Medline: 30924790]
16. Zheng X, Spatz ES, Bai X, et al. Effect of text messaging on risk factor management in patients with coronary heart disease: the CHAT randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019 Apr;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
17. Ershad Sarabi R, Sadoughi F, Jamshidi Orak R, Bahaadinbeigy K. The effectiveness of mobile phone text messaging in improving medication adherence for patients with chronic diseases: a systematic review. Iran Red Crescent Med J 2016;18(5):e25183. [doi: 10.5812/ircmj.25183] [Medline: 27437126]
18. DeKoekkoek T, Given B, Given CW, Ridenour K, Schueller M, Spoelstra SL. mHealth SMS text messaging interventions and to promote medication adherence: an integrative review. J Clin Nurs 2015 Oct;24(19-20):2722-2735. [doi: 10.1111/jocn.12918] [Medline: 26216256]

XSL•FO
RenderX

19. Thakkar J, Kurup R, Laba TL, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
20. Zhao YY, Dang FP, Zhai TT, Li HJ, Wang RJ, Ren JJ. The effect of text message reminders on medication adherence among patients with coronary heart disease: a systematic review and meta-analysis. Medicine (Baltimore) 2019 Dec;98(52):e18353. [doi: 10.1097/MD.0000000000018353] [Medline: 31876709]
21. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222. [doi: 10.2196/jmir.3442] [Medline: 25326646]
22. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15:33. [doi: 10.1186/s12913-014-0654-6] [Medline: 25609559]
23. Taylor D, Lunny C, Lolić P, et al. Effectiveness of text messaging interventions on prevention, detection, treatment, and knowledge outcomes for sexually transmitted infections (STIs)/HIV: a systematic review and meta-analysis. Syst Rev 2019 Jan 8;8(1):12. [doi: 10.1186/s13643-018-0921-4] [Medline: 30621784]
24. McLaughlin MM, Hsue PY, Lowe DA, Olgin JE, Beatty AL. Development of text messages for primary prevention of cardiovascular disease in persons with HIV. Cardiovasc Digit Health J 2023;4(6):191-197. [doi: 10.1016/j.cvdhj.2023.11.002] [Medline: 38222100]
25. Williams SK, Johnson BA, Tobin JN, et al. Protocol paper: stepped wedge cluster randomized trial translating the ABCS into optimizing cardiovascular care for people living with HIV. Prog Cardiovasc Dis 2020;63(2):125-133. [doi: 10.1016/j.pcad.2020.02.003] [Medline: 32035124]
26. Gamble-George JC, Longenecker CT, Webel AR, et al. ImPlementation REsearCh to DEvelop Interventions for People Living with HIV (the PRECluDE consortium): combatting chronic disease comorbidities in HIV populations through implementation research. Prog Cardiovasc Dis 2020;63(2):79-91. [doi: 10.1016/j.pcad.2020.03.006] [Medline: 32199901]
27. Lloyd-Jones DM, Huffman MD, Karmali KN, et al. Estimating longitudinal risks and benefits from cardiovascular preventive therapies among Medicare patients: the Million Hearts Longitudinal ASCVD Risk Assessment Tool: a special report from the American Heart Association and American College of Cardiology. J Am Coll Cardiol 2017 Mar 28;69(12):1617-1636. [doi: 10.1016/j.jacc.2016.10.018] [Medline: 27825770]
28. Frieden TR, Berwick DM. The “Million Hearts” initiative--preventing heart attacks and strokes. N Engl J Med 2011 Sep 29;365(13):e27. [doi: 10.1056/NEJMp1110421] [Medline: 21913835]
29. Suffoletto B. Deceptively simple yet profoundly impactful: text messaging interventions to support health. J Med Internet Res 2024 Aug 27;26:e58726. [doi: 10.2196/58726] [Medline: 39190427]
30. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026. [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
31. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. Med Decis Making 2007;27(5):672-680. [doi: 10.1177/0272989X07304449] [Medline: 17641137]
32. Matheny AP, Wachs TD, Ludwig JL, Phillips K. Bringing order out of chaos: psychometric characteristics of the Confusion, Hubbub, and Order Scale. J Appl Dev Psychol 1995 Jul;16(3):429-444. [doi: 10.1016/0193-3973(95)90028-4]
33. Fiscella K, Boyd M, Brown J, et al. Activation of persons living with HIV for treatment, the great study. BMC Public Health 2015 Oct 16;15:1056. [doi: 10.1186/s12889-015-2382-1] [Medline: 26474979]
34. Sanders M, Tobin JN, Cassells A, et al. Can a brief peer-led group training intervention improve health literacy in persons living with HIV? Results from a randomized controlled trial. Patient Educ Couns 2021 May;104(5):1176-1182. [doi: 10.1016/j.pec.2020.10.031] [Medline: 33221117]
35. Mattila E, Lappalainen R, Välkkynen P, et al. Usage and dose response of a mobile Acceptance and Commitment Therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90. [doi: 10.2196/mhealth.5241] [Medline: 27468653]
36. Li Y, Guo Y, Hong YA, et al. Dose-response effects of patient engagement on health outcomes in an mHealth intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jan 4;10(1):e25586. [doi: 10.2196/25586] [Medline: 34982724]
37. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019 Dec;38(12):1168-1174. [doi: 10.1037/hea0000805] [Medline: 31580127]
38. Bibbins-Domingo K, U.S. Preventive Services Task Force. Aspirin use for the primary prevention of cardiovascular disease and colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2016 Jun 21;164(12):836-845. [doi: 10.7326/M16-0577] [Medline: 27064677]
39. Brar Prayaga R, Agrawal R, Nguyen B, et al. Impact of social determinants of health and demographics on refill requests by Medicare patients using a conversational artificial intelligence text messaging solution: cross-sectional study. JMIR Mhealth Uhealth 2019 Nov 18;7(11):e15771. [doi: 10.2196/15771] [Medline: 31738170]
40. Rashid A, Zeb MA, Rashid A, Anwar S, Joaquim F, Halim Z. Conceptualization of smartphone usage and feature preferences among various demographics. Cluster Comput 2020 Sep;23(3):1855-1873. [doi: 10.1007/s10586-020-03061-x]

XSL•FO
RenderX

41. Read GL, Yan HY, Anderson PB, et al. Making stability dependable: stable cellphone access leads to better health outcomes for those experiencing poverty. Inf Commun Soc 2022;25(14):2122-2139. [doi: 10.1080/1369118X.2021.1928263]
42. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
43. Carroll JK, Tobin JN, Luque A, et al. "Get Ready and Empowered About Treatment" (GREAT) study: a pragmatic randomized controlled trial of activation in persons living with HIV. J Gen Intern Med 2019 Sep;34(9):1782-1789. [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
44. Ngiam NH, Yee WQ, Teo N, et al. Building digital literacy in older adults of low socioeconomic status in Singapore (Project Wire Up): nonrandomized controlled trial. J Med Internet Res 2022 Dec 2;24(12):e40341. [doi: 10.2196/40341] [Medline: 36459398]
45. Sparapani R, Spanbauer C, McCulloch R. Nonparametric machine learning and efficient computation with Bayesian additive regression trees: the BART R package. J Stat Softw 2021;97(1):1-66. [doi: 10.18637/jss.v097.i01]

Abbreviations

ABCS: aspirin therapy, blood pressure control, cholesterol management, and smoking cessation
ASCVD: atherosclerotic cardiovascular disease
BART: Bayesian additive regression trees
BAT: bidirectional automated texting
CHAOS: Confusion, Hubbub, and Order Scale
CVD: cardiovascular disease
FQHC: Federally Qualified Health Center
PAM: Patient Activation Measure
SNS: Subjective Numeracy Scale
ZINB: zero-inflated negative binomial

Edited by A Coristine; submitted 21.Nov.2024; peer-reviewed by A Bermon, S Ekongefeyin; revised version received 27.Apr.2026; accepted 30.Apr.2026; published 14.Jul.2026.

Please cite as:
Sanders M, Harrington D, Sass E, Thomas M, Holder T, Tian Y, Johnson B, Cassells A, Tobin JN, Fiscella K
Bidirectional Automated Texting for Cardiovascular Health Among People Living With HIV: Observational Cohort Analysis of a Stepped-Wedge Cluster Randomized Controlled Trial
JMIR Cardio 2026;10:e69098
URL: https://cardio.jmir.org/2026/1/e69098
doi:10.2196/69098

© Mechelle Sanders, Donald Harrington, Emma Sass, Marie Thomas, Tameir Holder, Yiqi Tian, Brent Johnson, Andrea Cassells, Jonathan N Tobin, Kevin Fiscella. Originally published in JMIR Cardio (https://cardio.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mindfulness-Based Self-Management Program Using a Mobile App for Patients With Pulmonary Hypertension: Single-Arm Feasibility Study

Yuka Takita[1], PhD; Junko Morishita[2], PhD; Sunre Park[3], MA; Ayumi Goda[4], MD, PhD; Takumi Inami[4], MD, PhD; Hanako Kikuchi[4], MD; Takashi Kohno[4], MD, PhD; Masaharu Kataoka[5], MD, PhD; Daisuke Fujisawa[6], MD, PhD

[1]Department of Nursing, Faculty of Medical Technology, Teikyo University, 2-11-1, Kaga, Itabashi-ku, Tokyo, Japan

[2]Department of Nursing, National College of Nursing, Japan Institute for Health Security, Tokyo, Japan

[3]Department of Nursing, Faculty of Nursing and Medical Care, Keio University, Tokyo, Japan

[4]Department of Cardiovascular Medicine, Faculty of Medicine, Kyorin University, Tokyo, Japan

[5]The Second Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan

[6]Division of Quality Assurance Programs, National Cancer Center Institute for Cancer Control, Tokyo, Japan

Corresponding Author:
Yuka Takita, PhD
Department of Nursing, Faculty of Medical Technology, Teikyo University, 2-11-1, Kaga, Itabashi-ku, Tokyo, Japan

Abstract

Background: Mindfulness-based interventions have been applied across various chronic illnesses, but no tailored program exists for individuals with pulmonary hypertension (PH).

Objective: This study aimed to develop and evaluate the feasibility of a mindfulness-based self-management program for patients with PH, delivered online to accommodate their limited mobility.

Methods: A single-arm pre-post study was conducted using an 8-session, weekly videoconference program incorporating PH self-management education and elements of mindfulness-based cognitive therapy. A mobile app linked to an Apple Watch was used to support symptom monitoring and mindfulness awareness. Outcomes included PH-related symptoms, quality of life (emPHasis-10), depression (Patient Health Questionnaire-9 [PHQ-9]), anxiety (Generalized Anxiety Disorder 7-item scale [GAD-7]), resilience (Connor-Davidson Resilience Scale [CD-RISC]), and loneliness (UCLA Loneliness Scale–short version). Assessments occurred at baseline, week 4, and program completion. Exit interviews explored perceived changes and experiences.

Results: Twelve participants (mean age 41.8, SD 10.5 years; range 26 - 56 years) were enrolled, and 9 completed the program (75% retention). Participants valued the online format and Apple Watch integration, while noting a need for optional on-demand sessions. Qualitative analysis identified themes such as increased self-awareness, use of meditation for pain management, and enhanced self-compassion. Quantitative analysis showed significant changes across 3 time points (baseline, week 4, and week 8) for emPHasis-10 (χ^2_2 =9.74; *P*=.008) and CD-RISC (χ^2_2 =7.27; *P*=.03). Trends toward change were observed for PHQ-9 (χ^2_2 =4.75; *P*=.09) and GAD-7 (χ^2_2 =5.07; *P*=.08), but week 12 data were limited (n=5). No significant changes in loneliness were observed.

Conclusions: The program appeared to support patients with PH in managing symptoms and emotions and suggested potential improvements in quality of life. These preliminary findings warrant evaluation in a future randomized controlled trial.

Trial Registration: UMIN Clinical Trials Registry UMIN000044075; https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000050319

(JMIR Cardio 2026;10:e79639) doi:10.2196/79639

KEYWORDS

mindfulness; mindfulness-based intervention; pulmonary hypertension; self-management; digital health

Introduction

Pulmonary hypertension (PH) is a progressive disease characterized by shortness of breath as the primary symptom. Although recent advances in treatment have dramatically improved the prognosis of the disease [1], patients with PH may face various physical symptoms and may experience limitations in their activities and social roles. Self-management at home has become increasingly complex owing to the use of medications with various mechanisms of action and with various routes of administration [2]. Furthermore, side effects of

XSL•FO
RenderX

pulmonary vasodilative medications, such as headache, jaw pain, plantar pain, diarrhea, and nausea, may occur. These symptoms are more pronounced in continuous intravenous and subcutaneous infusion therapies. There is a certain level of risk for sudden death in severe PH cases with right heart failure; thus, patients with PH are forced to live with uncertainty.

Therefore, it is not surprising that patients with PH tend to experience high anxiety and depression. Past studies demonstrated that 21.9% - 56% of PH patients are comorbid with depressive symptoms [3-14], 10.7% - 62% with anxiety symptoms [3-6,8,12,13], and 27.6% - 40% with stress-related symptoms [4,11]. These findings suggest the need for psychological care for patients with PH.

Mindfulness has been described as "paying attention in a particular way—on purpose, in the present moment, and nonjudgmentally" [15]. Mindfulness-based interventions have gained increasing attention as a method of psychological care for patients with physical illnesses. The two major programs are mindfulness-based stress reduction (MBSR) [16] and mindfulness-based cognitive therapy (MBCT) [17]; however, flexible modifications of the program to fit with target populations, such as providing information on nutrition [18], grief care [19], and advance care planning [20,21], have been implemented.

Conventional mindfulness-based programs have been delivered on a face-to-face basis; however, in recent years, web-based programs have been developed and implemented quite widely [22,23]. In the field of cardiovascular disease, mindfulness interventions are being developed for patients with coronary artery disease and heart failure [24-26].

However, to the best of the authors' knowledge, there has been no mindfulness-based program that has been developed specifically for PH. Modification of the program may be needed for patients with PH to avoid physical overload that may worsen right ventricular function. For example, yoga, a standard component of MBSR and MBCT, could lead to cardiopulmonary overload. Furthermore, since patients with PH often have limited mobility, web-based programs, instead of on-site programs, are likely to be preferred.

Therefore, in this study, we developed and tested the feasibility of an online mindfulness-based self-management program for patients with PH using a smartphone app on Apple Watch. The intervention aimed to improve the quality of life (QOL) and resilience and to reduce depression, anxiety, and pain (a side effect of the treatment) of patients with PH.

Methods

Study Design

This study used a mixed methods, single-group, pre-post design.

Participants

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria are shown in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

The inclusion criteria for this study were as follows:

- A confirmed diagnosis of pulmonary hypertension (PH) [27].
- Age 20 - 75 years.
- The ability to attend at least six of the eight 60-minute online sessions.
- The ability to operate the self-administered mobile app on an iPhone or iPad and an Apple Watch.

The exclusion criteria were as follows:

- Severe physical symptoms, including but not limited to decompensated right-sided heart failure, as assessed by the attending physician.
- Patients with an active psychiatric disorder who were currently under psychiatric care or receiving psychotropic treatment. These individuals were excluded for safety reasons, as the program was delivered entirely online, and timely in-person support could not be ensured if psychological distress or other adverse emotional reactions were to occur during the intervention.
- Cognitive impairment or other conditions that, in the attending physician's judgment, would make it difficult for the patient to understand or participate in the program, based on routine clinical assessments and information in medical records.
- Individuals deemed by their attending physicians to be clinically unstable or otherwise inappropriate for participation.
- Individuals who had previously participated in structured mindfulness-based programs such as mindfulness-based cognitive therapy (MBCT) or mindfulness-based stress reduction (MBSR).

Individuals aged 18 - 19 years were not included, because at the time this study was initiated (December 2020), persons ages 20 years and younger were not considered adults for the purpose of providing independent informed consent under Japanese ethical guidelines, and thus, were unable to consent without guardian approval [28].

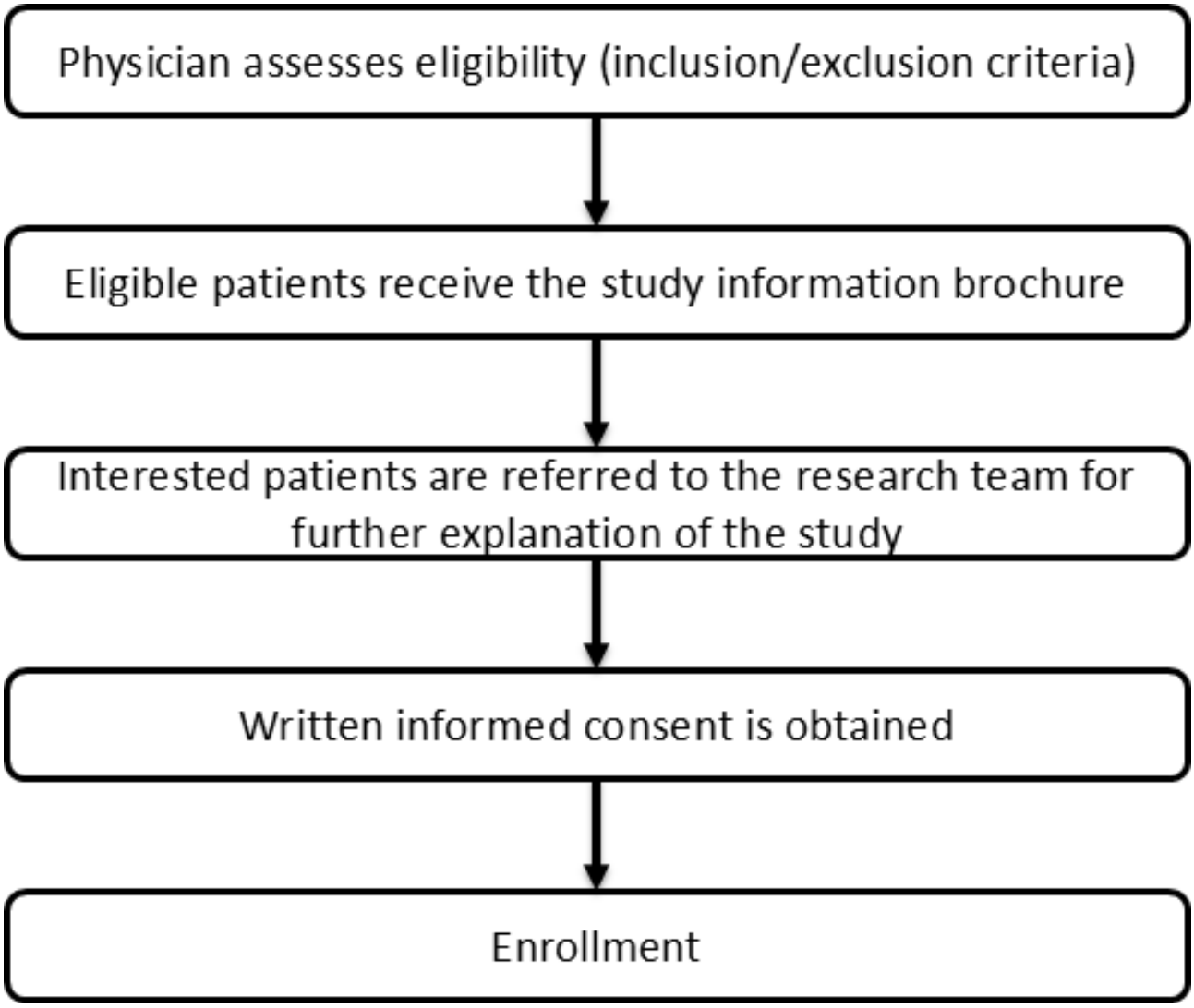
Recruitment and Screening Procedures

Participants were recruited through a structured, stepwise process. First, physicians screened patients during routine clinical visits to ensure they met the inclusion criteria and did not meet any exclusion criteria. Eligible patients were then provided with an informational brochure describing the study. Patients who expressed interest were referred to the research team, who provided a detailed explanation of the study

XSL•FO
RenderX

procedures. Written informed consent was obtained from all patients before enrollment. The overall recruitment and screening process is shown in Figure 1.

Figure 1. Flowchart illustrating the recruitment and screening.



Sample Size Justification

As this was a feasibility study, the primary aim was to evaluate acceptability, adherence, and operational feasibility rather than to test statistical efficacy; therefore, a formal power calculation was not required. We chose a target sample size of approximately 12 participants, which is comparable to previous feasibility studies of meditation or mindfulness-based interventions in cardiovascular populations that enrolled about 10 - 15 participants [29,30].

Intervention

Program Contents

The conceptual framework and content of the program are presented in Figure 2 and Table 1. We developed this program based on the findings of our previous research, which showed that the key elements of distress for patients with PH were the loss of the past and the threat of disease progression, including rumination about the past and concern about the future [14]. The fundamentals of the intervention were based on MBCT, which addresses rumination and has been proven effective for both physical and psychological symptoms in patients with critical illnesses [17]. Furthermore, we added psychoeducation and self-management skill-building as essential components, as the distress of patients with PH derives from PH-specific symptoms (eg, breathlessness) and the adverse effects (eg, pain or nausea) and difficulties associated with the treatment (eg, home oxygen therapy).

XSL•FO
RenderX

Figure 2. Conceptual framework of the mindfulness-based intervention program for patients with pulmonary hypertension.

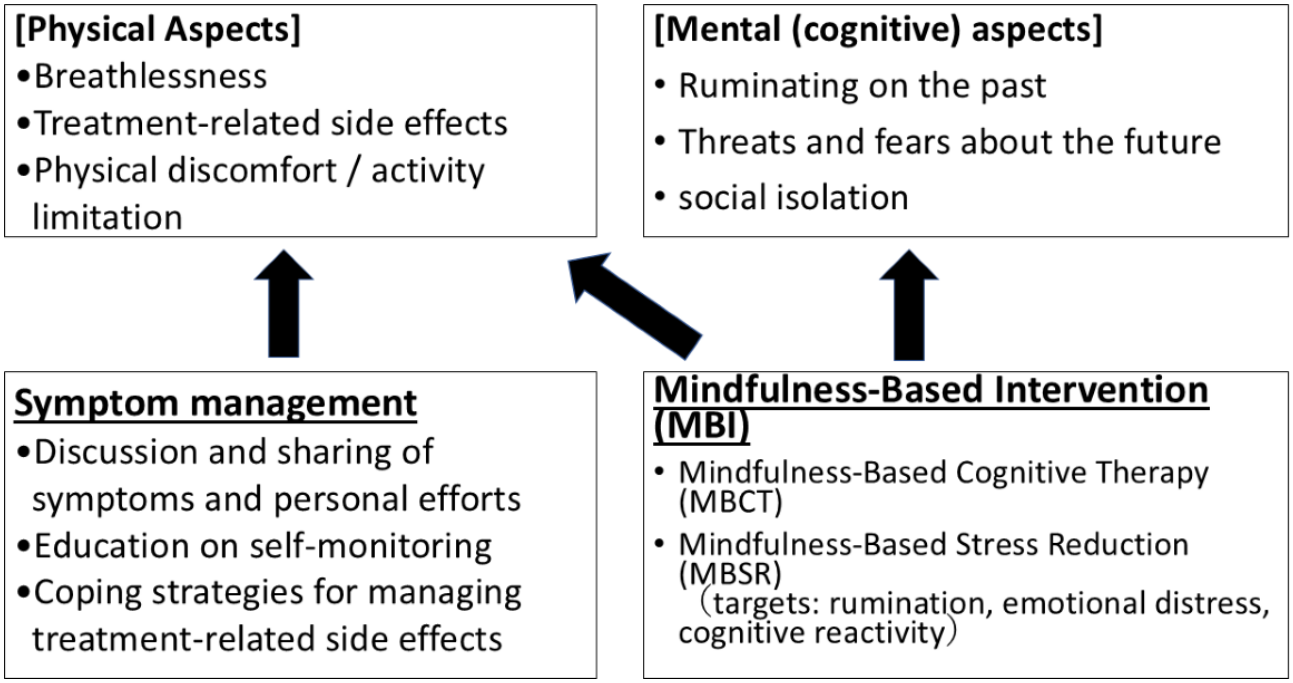


Table . Program content structured based on the results of previous research.

Elements of distress	Program content and expected effects
“Isolation from my surroundings”	Prevent isolation and reduce loneliness by regularly connecting with others and sharing thoughts and feelings through weekly online meetings.
“Loss of myself” and “Fear of illness progression or deterioration”	By reflecting on their physical and mental conditions through mindfulness practice and homework, participants become aware of their thought patterns and feelings, such as loss of abilities, regret regarding the past, and anxiety about the future; promoting meta-cognition stops rumination, thereby alleviating mood swings and anxiety.
“Hassle associated with oxygen therapy,” “Suffering from side effects,” and “Rumination on illness due to breathlessness”	Learning and sharing the basic knowledge and practical tips required for self-management and enhancing the ability to self-manage activities, shortness of breath, side effects, and other issues.

Program Structure

Table 2 presents the outline of the program. The basic structure of the program was similar to that of MBCT. To lessen the potential physical and psychological burden of the patients, we adapted an online format. We shortened the length of each session to 1 hour. We eliminated yoga, which has been supposed to be an integral component in conventional MBCT, for safety reasons, since patients with PH are at risk of circulatory collapse owing to direct right ventricular stress caused by increased pulmonary artery pressure, which could be exacerbated by physical exercise.

XSL•FO
RenderX

Table . Schedule and homework for the self-management mindfulness program for patients with pulmonary hypertension.

Session	Theme	Contents	Homework
1	Start the program	• What is mindfulness? • How to use the self-management app? • Short breathing meditation	• Using self-management app
2	Become aware of automatic reactions	• Short breathing meditation • Symptom management of pulmonary hypertension	• Mindful daily living
3	Focus on your body	• Body scan (meditation) • Cognitive strategies for managing treatment-related side effects	• Body scan
4	Focus on breathing	• Mindfulness meditation of breathing and body • Pleasant and unpleasant mode	• Pleasant and unpleasant life diary
5	Focus on your body and adjust it	• Short meditation • Adjustment of pulmonary hypertension activities	• Meditation • Focus on your pulse during activity
6	Thoughts are not facts	• Mindfulness meditation with sound and thought • Cognitive strategies for managing thoughts and emotions (guided imagery practice) • 3-step breathing space method	• Using 3-step breathing space method when feeling uncomfortable
7	Caring for yourself	• Compassion meditation • Positive habits	• Meditation • Appreciation list
8	Use skills for your future	• Meditation • Reflection on the past • Looking toward the future	• —[a]

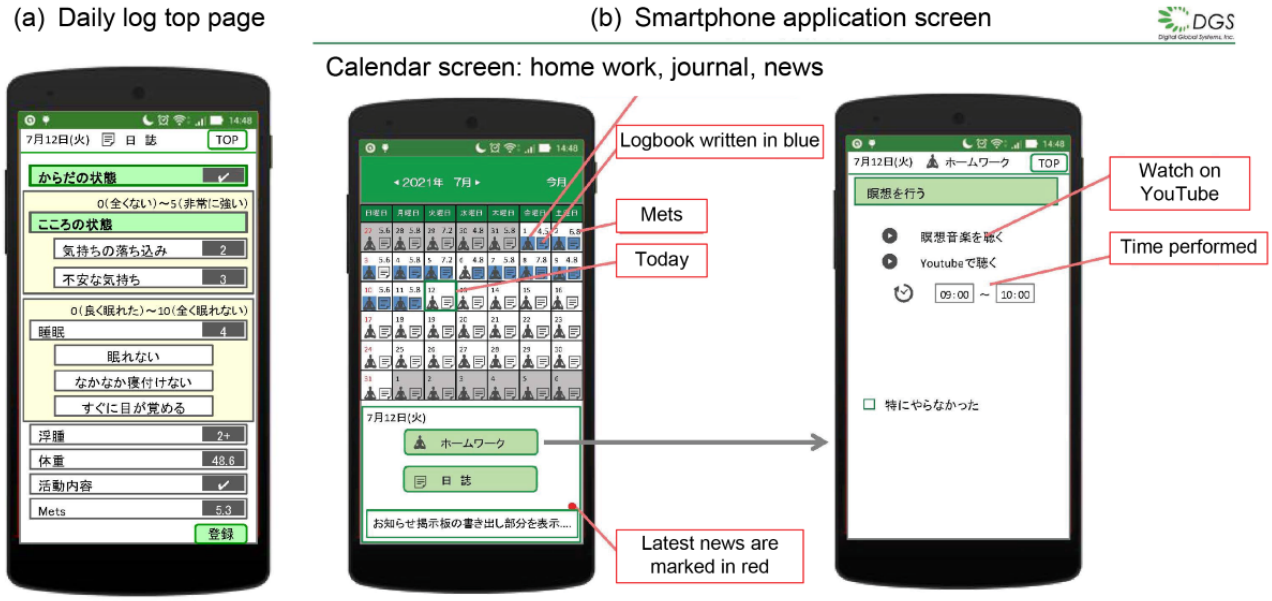
[a]Not applicable.

The program was delivered in a group format once a week over a period of 8 weeks via video conference. The program was facilitated by the lead author, a nurse with 6 years of experience in practice and the teaching of mindfulness interventions. Each session lasted 60 minutes, which included 10- to 30-minute meditation practice, which was facilitated with the facilitator. At the beginning of each session, a reflection on homework was conducted, and an inquiry session was held after each meditation practice. Homework included daily mindfulness practice (eg, breathing meditation and body-scan meditation) and reflection journaling on emotional and physical experiences. Video lectures on the meaning of mindfulness, the pleasant and unpleasant mode, and mindfulness meditation were provided by a clinical psychologist and a nurse who had been a mindfulness provider for over 10 years. Pleasant mode and unpleasant mode involve mindfully reflecting on pleasant and unpleasant experiences in daily life, noting the physical sensations, emotions, and thoughts that arose at the time, thereby cultivating awareness.

To facilitate self-management skills, we developed a mobile app, which enabled users to record their daily physical conditions and to monitor their activity status using an Apple Watch (Figure 3). The mobile app (Self-management App, developed in collaboration with DGS Co Ltd) stored data on secure encrypted servers; all identifying information was anonymized and replaced by unique participant codes. For participants who did not own an Apple Watch, a device was lent to them free of charge for use during the study period. Participants were advised to use this mobile app throughout the program.

XSL•FO
RenderX

Figure 3. Screenshots of the self-management smartphone app used in the study.



Outcome Measures

Demographic Data

Demographic and clinical information, including age, sex, diagnosis, disease classification (World Health Organization functional class), duration of illness, and treatment status, were collected at baseline through participants' self-report and review of medical records. Demographic information was obtained to characterize the study sample and was used only for descriptive analyses.

Feasibility and Acceptability

The feasibility of the intervention was evaluated with participation and completion rates for the program. Acceptability was assessed by collecting and evaluating feedback from the program through interviews. The interviews included questions about changes in symptoms, self-management skills, and mental health conditions before and after the program. Furthermore, the participants were asked to describe the perceived benefits of the program, what should be changed, what they liked and disliked, and obstacles and facilitators for implementation of the program.

Secondary Outcomes

Secondary outcome measures comprised emPHasis-10, a disease-specific patient-reported outcome measure for evaluating the QOL of patients with PH [31]; the Patient Health Questionnaire (PHQ-9) for depression [32]; the Generalized Anxiety Disorder 7-item (GAD-7) scale [33]; the Connor-Davidson Resilience Scale (CD-RISC), a measure of one's ability to recover from various difficulties such as illness, emotional pressure, and pain [34]; and the UCLA Loneliness Scale-short version [35]. In addition, interviews were conducted regarding changes in awareness and behaviors related to self-management. These measures were obtained at baseline, 4 weeks after the program, and at the end of the program. The questionnaire was administered at four time points (baseline, week 4, week 8, and week 12). All questionnaires were provided in paper format, completed by participants at each time point, and returned by mail.

Analytical Methods

Content analysis was used to analyze the qualitative data from the interviews [36]. The characteristics and results of the participants were summarized using descriptive statistics, with median (range) for continuous variables and frequency (percentage) for categorical variables. The score changes between time points for each participant were plotted as line graphs for each scale. The Friedman test was conducted to evaluate the differences in scores across time points. This nonparametric test was chosen owing to its suitability for comparing repeated measures or related samples without assuming normality. Significance was set at $P<.05$. IBM SPSS (version 28) for Windows was used for all statistical analyses. MAXQDA (VERBI Software GmbH) was used for qualitative data analysis.

Ethical Considerations

This study was approved by the ethics review committees of Tokyo Kasei University and Kyorin University School of Medicine (SKE2020-11, number 1631) and was conducted in accordance with the ethical standards of the Declaration of Helsinki and relevant national guidelines.

All participants received a written explanation of the study's purpose, procedures, potential risks, and their right to withdraw at any time without disadvantage. Written informed consent was obtained from all participants prior to study participation.

To ensure privacy and confidentiality, all data were anonymized and assigned participant identification codes at the time of collection. No personally identifiable information was included in the datasets used for analysis. Data were securely stored on a password-protected hard disk drive.

Participants did not receive monetary compensation but were provided with the mindfulness program free of charge as part of the study.

XSL•FO
RenderX

No identifiable images or personal information of participants are included in this paper or the supplementary files.

The study was registered in the UMIN Clinical Trials Registry (UMIN000044075).

Results

Participant Characteristics

Participants' baseline characteristics are shown in Table 3.

XSL•FO
RenderX

Table . Participant characteristics at baseline (n=12).

Characteristics	Results
Age (years), mean (SD)	41.8 (10.5)
Age (years), range	26 - 56
Sex, n (%)	
Male	3 (25)
Female	9 (75)
Program completed, n (%)	9 (75)
PH[a] type, n (%)	
PAH[b]	11 (92)
CTEPH[c]	1 (8)
MeanPAP[d] (mmHg), median (range)	32.5 (20-47)
Treatment (pharmacological therapy), n (%)	
Epoprostenol (IV[e])	5 (42)
Treprostinil (IV)	2 (2)
Treprostinil (SC[f])	1 (8)
Only oral medicine[g]	4 (33)
Interventional history, n (%)	
Post BPA[h]	1 (8)
Therapeutic support, n (%)	
Oxygen therapy	3 (25)
Symptoms, n (%)	
Dyspnea on exertion	10 (83)
Fatigue	10 (83)
Palpitations	3 (25)
Pain	8 (67)
Nausea	3 (25)
Diarrhea	5 (42)
QOL[i] (emPHasis-10[j]), mean (SD), range	33.6 (10.1), 16 - 46
Depression (PHQ-9[k]), mean (SD), range	14.1 (6.6), 2 - 24
Anxiety (GAD-7[l]), mean (SD), range	11.1 (5.5), 1 - 18
Resilience (CD-RISC[m]), mean (SD), range	17.8 (8.0), 3 - 29
Loneliness (UCLA Loneliness Scale), mean (SD), range	26.6 (5.9), 13 - 34

[a]PH: pulmonary hypertension.
[b]PAH: pulmonary arterial hypertension.
[c]CTEPH: chronic thromboembolic pulmonary hypertension.
[d]MeanPAP: mean pulmonary arterial pressure.
[e]iv: intravenous injection therapy.
[f]sc: subcutaneous injection therapy.
[g]Including the 1 participant with post-BPA status.
[h]BPA: balloon pulmonary angioplasty (only applies to patients with CTEPH).
[i]QOL: quality of life.
[j]emPHasis-10: disease-specific patient-reported outcome (PRO) measures for evaluating quality of life in patients with pulmonary hypertension.
[k]PHQ-9: Patient Health Questionnaire-9.

XSL•FO
RenderX

[l]GAD-7: Generalized Anxiety Disorder 7-item scale.

[m]CD-RISC: Connor-Davidson Resilience Scale.

Feasibility

Of the 12 participants who agreed to participate in this study, 9 completed the program (75% retention rate). The 3 individuals dropped out due to hospitalization for exacerbation of hemoptysis, hospitalization for treatment of comorbidities, and catheter infection.

Acceptability of Program Structure

The responses by the 9 participants who completed the program are shown in Table 4. Furthermore, the following comments were received regarding the program's structure:

> *The weekly online real-time program made me feel as though I were back at work again, and it refreshed me and made me feel lively and excited.*
>
> *The Apple Watch helped me understand my physical condition better through objective numbers and data, and it gave me the opportunity to face my illness properly. Just experiencing symptoms can be vague, and they are easily forgotten, but being able to check the numbers and monitor my condition is really helpful.*
>
> *I didn't know anything about mindfulness at first, so I felt hesitant. I think it would be much easier to engage if I could learn why mindfulness is applied to this illness.*
>
> *I take diuretics every day, so it would be helpful if there were an on-demand option that allows me to participate calmly without worrying about needing to urinate.*

Table . Feedback on the number of sessions, duration per session, and session intervals (n=9).

Variable	Sample, n (%)
Number of programs (8 in total)	
Too many	2 (22)
Appropriate	6 (67)
Too few	1 (11)
Time per session (60 minutes per session)	
Too long	0 (0)
Appropriate	9 (100)
Too short	0 (0)
Program interval (once per week)	
Too long	0 (0)
Appropriate	7 (78)
Too short	2 (22)

Changes and Awareness After the Program

The qualitative content analysis identified 10 categories (Table 5).

XSL•FO
RenderX

Table . Qualitative analysis results: categories reflecting changes and awareness after the program.

Category	Frequency, n
Being able to objectively consider one's thoughts and feelings and not be bothered by them	14
Peace of mind from facing one's mind and body	13
Reduced pain and handling side effects	12
Being compassionate toward oneself	10
Focusing on one's body and conducting activities	10
Awareness of the importance of breathing	6
Increased happiness and positive feelings	4
Realization that actions during illness until the present were not erroneous	4
Reduced fatigue	3
Acquiring methods of thinking and coping when feeling anxious or restless	2

Being Able to Objectively Consider One's Thoughts and Feelings and Not Be Bothered by Them

Participants were able to understand that their mental instability was due to their poor physical condition.

> *I used to struggle to accept it, but now I can think, 'My physical condition is bad today, so it's natural [for my feelings to also decline]. It can't be helped.' […] And I can tell myself, 'It's the illness that makes me feel this way, so it can't be helped.'* [ID7]

Reduced Pain and Handling Side Effects

They experienced how meditation can help relieve pain and heartburn.

> *I had been having a headache for a long time, but while I was doing the body scan meditation, the pain disappeared… It's really strange, but the pain seems to move around. I always wonder why.* [ID5]

Being Compassionate Toward Oneself

Even when thoughts and feelings came to mind, they were able to observe them objectively and respond more adaptively. Consequently, they felt that this positive change reduced self-blame and harsh self-criticism.

> *I really felt that the number of times I thought, 'Oh no, this is terrible,' had decreased...I really felt that the number of times I thought.* [ID2]

Focusing on One's Body and Conducting Activities

By focusing on bodily sensations and intentionally attending to their internal states during daily activities, participants reported reduced pain.

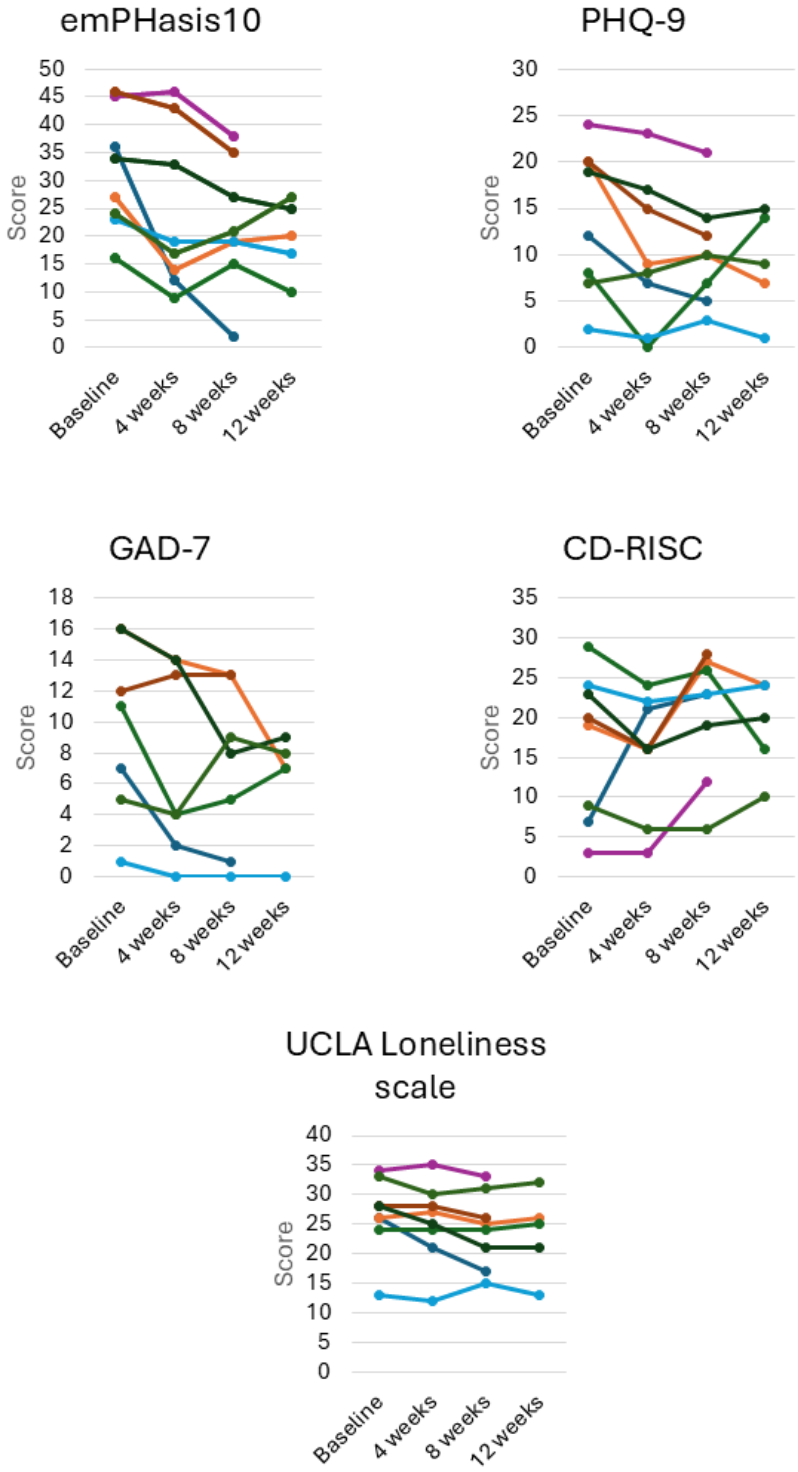
> *At night, after eating dinner and taking my medication, my heart used to start pounding faster. But now, after I finish eating, I take a moment and tell myself, 'Let's breathe slowly and take a little rest,' and the breathlessness goes away.* [ID4]

Secondary Outcomes

Although 9 participants completed the program (75% completion rate), the week 8 quantitative analysis—the primary endpoint—included data from 8 participants (67% data retention rate) because one completer did not return the postprogram questionnaire. Figure 4 shows score trends from baseline, week 4, week 8, and week 12. At week 12 (one month after the end of the program), responses were obtained from 5 participants (56% response rate).

XSL•FO
RenderX

Figure 4. Individual trajectories of patient-reported outcome measures across the four assessment time points (baseline, 4 weeks, 8 weeks, and 12 weeks). The figure displays individual-level changes in 5 psychosocial and clinical measures used in this study: the emPHasis-10 (disease-specific quality of life), PHQ-9 (depressive symptoms), GAD-7 (anxiety symptoms), CD-RISC (resilience), and the UCLA Loneliness Scale (loneliness). Each colored line represents a single participant's score over time. These descriptive data illustrate variability in symptom patterns and potential trends in psychological well-being during and after the mindfulness-based intervention. CD-RISC: Connor-Davidson Resilience Scale; GAD-7: Generalized Anxiety Disorder 7-item scale; PHQ-9: Patient Health Questionnaire-9.



The emPHasis-10 (χ^2_2=9.74; *P*=.008) and CD-RISC (χ^2_2=7.27; *P*=.03) scores showed significant differences at baseline (preprogram), 4 weeks (during the program), and 8 weeks (postprogram; n=8). The significant differences were not maintained at 12 weeks (follow-up). In addition, although no significant difference was found, a trend toward improvement in the PHQ-9 (χ^2_2=4.75; *P*=.09) and GAD-7 (χ^2_2=5.07; *P*=.08) was observed. Moreover, a few participants showed sudden deterioration in QOL and experienced depressive and anxiety symptoms a month after the program; however, they reported

XSL•FO
RenderX

that this coincided with the timing of replacing the subcutaneous Treprostinil injection needle. No significant changes were observed in feelings of loneliness. Individual score trajectories are shown in Figure 4.

Discussion

Principal Findings

This study evaluated the feasibility, accessibility, and preliminary psychological changes associated with an online mindfulness-based program designed for patients with PH. The program achieved good retention, with a completion rate of 75% (9/12), and was generally well accepted by participants. Quantitative analyses revealed positive changes in QOL, depression, anxiety, and resilience scores, while qualitative findings indicated that participants experienced greater bodily awareness, emotional regulation, and self-compassion through mindfulness practice. Together, these results suggest that an online mindfulness-based intervention is both feasible and acceptable for patients with PH and may have the potential to improve their psychological well-being and symptom management.

Feasibility and Accessibility

With a completion rate of 75% (9/12), the program demonstrated a relatively high level of retention. However, the primary reason for participant dropout was hospitalization due to physical deterioration associated with underlying or comorbid conditions. As PH progresses, patients may develop worsening right heart failure or experience hemoptysis, necessitating hospitalization. Therefore, future implementations of the program should target individuals with stable PH and take into account the potential for physical decline during participation.

The 60-minute session length appeared appropriate; however, participants indicated that the total number of sessions (n=8) was excessive and that the weekly interval was too short. Additionally, some participants expressed a desire for on-demand sessions, citing difficulties caused by frequent urination due to diuretic use. These findings suggest the need to reconsider the program's frequency and delivery format.

The program consisted of eight 60-minute real-time sessions conducted weekly via an online conferencing system. For PH patients who experience challenges leaving their homes, this online delivery format effectively reduced their physical burden. Nevertheless, the requirement to connect online at a fixed time each week imposed an additional burden that may have negatively affected retention. Conversely, participants who had stopped working due to illness reported a sense of fulfillment and accomplishment from attending regularly scheduled sessions.

As discussed above, offering the entire program on an on-demand basis could further reduce participant burden and improve retention. However, inquiry—a central element of mindfulness—requires real-time interaction between instructors and participants. Inquiry, also referred to as mindful dialog, provides participants with opportunities for self-reflection through guided conversation. To preserve this essential component, at least part of the program should be conducted synchronously rather than fully on-demand. Although traditional MBSR and MBCT programs comprise 8 sessions, recent studies have developed and validated shorter mindfulness-based interventions [37-40]. Based on these findings, we propose a hybrid format that reduces the total number of sessions and delivers approximately half of them on demand. This approach may lessen participant burden while maintaining the feasibility and therapeutic integrity of the intervention. Nevertheless, some patients may require individualized support to manage specific physical symptoms, treatment-related side effects, or emotional distress. Therefore, it may be necessary to consider implementing individual consultations before the program or during the intervention period to provide tailored support for such patients.

Furthermore, this intervention was implemented as an individual program. Expanding individual programs to a larger population poses logistical challenges and limits generalizability. Future studies should therefore examine the effectiveness of group-based formats. In addition, since mindfulness interventions require trained facilitators, the current shortage of qualified personnel represents a barrier to broader dissemination. To promote scalability, two approaches may be considered: (1) developing on-demand mindfulness modules that can be facilitated by health care professionals without formal mindfulness training, or (2) establishing training programs to cultivate health care providers capable of delivering mindfulness interventions.

Although the response rates for pre- and midprogram questionnaires were high, those at program completion and 4 weeks postcompletion were notably low. In this study, paper-based questionnaires were distributed and returned by mail. For PH patients who find it burdensome to leave home, this method likely contributed to the low response rate. Therefore, future studies should employ web-based questionnaires that allow participants to respond online, thereby reducing the response burden and improving data collection rates.

Qualitative Findings: Psychological and Physical Experiences During the Program

Qualitative analysis identified themes such as "finding peace of mind through connecting with the body and mind" and "recognizing the importance of breathing." Interoception refers to perceiving, accessing, and evaluating internal bodily signals [38]. In mindfulness meditation, participants focus nonjudgmentally on present-moment bodily sensations. Rumination arises when attention drifts toward the past or future; mindfulness meditation helps disengage from rumination by redirecting awareness to the present body and activating interoceptive processing [41]. In this program, body-scan and breath-focused meditations enhanced participants' interoceptive awareness, helping them notice their breathing patterns, bodily reactions during activities, and the thoughts and emotions arising in their minds. Through nonjudgmental observation, participants experienced decentering, which enabled them to view their situations from a distance, thereby reducing rumination and perceived pain.

XSL•FO
RenderX

In addition, lectures and cognitive-behavioral activities addressing emotions and physical symptoms related to treatment side effects provided opportunities to learn coping strategies and methods for engaging with these symptoms. By increasing bodily awareness and practicing mindful engagement during activities, participants reported decreased fatigue and pain. These observations suggest that, in addition to mindfulness practice itself, learning management strategies tailored to PH-related side effects and adaptive coping methods may contribute to reducing treatment-related discomfort and improving activity tolerance.

Breathing was a key element of mindfulness meditation. Before the study, there was concern that focusing on breathing might exacerbate dyspnea in PH patients. Contrary to expectations, meditation focusing on breathing appeared to lessen perceived breathlessness. For individuals experiencing daily dyspnea, increasing awareness of breathing helped them recognize shallow breathing patterns and intentionally breathe more deeply and slowly, which may have contributed to reduced shortness of breath during exertion.

Furthermore, improvements in self-compassion, well-being, and positive emotions were observed after program participation. Many PH patients tend to blame themselves, thinking, “If only I had sought treatment earlier,” or “I became sick because I was weak.” Mindfulness may have helped reduce self-critical thinking by fostering objective awareness of mental states and thought patterns, encouraging nonjudgmental observation, and supporting more positive self-recognition. Overall, these findings suggest that the program may have a favorable influence on mental health and emotional well-being, though further investigation in larger, controlled studies is needed.

Comparison With Previous Work

Changes observed in QOL, depression, anxiety, and resilience scores suggest that the intervention may have the potential to support improvements in psychological well-being. However, immediate effects on rapidly worsening physical symptoms should not be expected. In this study, standardized self-administered questionnaires were conducted at four time points (baseline, week 4, week 8, and week 12). Because this was a feasibility study with a small sample size and a low response rate at week 12, statistical analyses across all time points were limited. Analyses of the first three time points (baseline, week 4, and week 8) indicated preliminary improvements in QOL, with several participants demonstrating downward trends in depression and anxiety scores. While these findings are not sufficient to determine efficacy, they suggest possible patterns of change that merit further investigation in future controlled trials.

Previous studies of mindfulness-based interventions in cardiovascular populations provide context for interpreting these findings. A randomized controlled trial (The Stress Reduction, Meditation, and Mindfulness Program) targeting patients with chronic heart failure reported reductions in perceived stress and improvements in clinical outcomes [26]. Similarly, a meta-analysis of randomized controlled trials in patients with coronary artery disease demonstrated significant reductions in depression, anxiety, and stress following mindfulness-based interventions [25]. Although the present study focused on patients with PH, the observed trends—increased emotional regulation, decreased rumination, and enhanced self-compassion—are consistent with mechanisms reported in these previous studies.

The mean CD-RISC score in this sample was lower than that reported in patients with cancer (29.3) [42] or multiple sclerosis (26.8) [43], suggesting relatively low baseline resilience in participants. While the small sample size limits interpretation, the observed changes in resilience scores suggest that mindfulness-based programs may hold potential for supporting resilience in PH populations. These preliminary findings align with previous literature indicating that mindfulness practice may contribute to improvements in adaptive coping and emotional well-being across diverse chronic illness populations.

Limitations

This study has several limitations that should be acknowledged.

First, as a single-arm feasibility study, it did not include a control group. Without a comparison to usual care or another intervention, it is difficult to determine whether the observed improvements in quality of life or psychological measures were attributable to the program itself or to other factors, such as natural adaptation or social interaction with facilitators. To mitigate this, standardized self-report measures were used at multiple time points to examine within-participant changes. Future studies should include a randomized controlled design to allow more robust evaluation of efficacy.

Second, the sample size was small, which limits the statistical power and generalizability of the findings. Because of the limited number of participants, subgroup analyses could not be performed, and some trends may have gone undetected. Nevertheless, the study provided valuable preliminary data on feasibility and participant experiences, which will help inform sample size estimation and stratification criteria for future trials.

Third, the study duration was relatively short, and no long-term follow-up was conducted. Consequently, it remains unclear whether the improvements observed immediately after the intervention can be sustained over time. Future studies should incorporate follow-up assessments to examine the persistence of psychological and physical benefits and to identify factors that influence continued engagement with mindfulness practice.

Fourth, response rates for the postprogram and 4-week follow-up questionnaires were low, possibly because of the paper-based mailing method, which may have introduced response bias. Participants who continued responding may have been those more motivated or satisfied with the program. To improve response rates and minimize bias, future research should use web-based data collection methods to facilitate participation and reduce the burden on patients with PH.

Future Directions

Based on these findings, several future directions for research and program development can be proposed. First, future studies should conduct randomized controlled trials with larger and more diverse samples to verify the program’s efficacy and assess its long-term impact. Longitudinal follow-up is necessary to

XSL•FO
RenderX

evaluate whether the psychological and physical benefits observed in this feasibility study can be sustained over time. Multisite or decentralized clinical trial designs may also facilitate participation among patients with PH who face difficulties traveling to research facilities.

Second, refinement of the program content is warranted. Adjustments to session frequency and duration, as well as partial incorporation of on-demand components, may help reduce participant burden and improve accessibility. Developing a group-based version of the program could further enhance social connectedness and scalability.

Third, to promote broader dissemination and generalization of mindfulness-based interventions, it will be important to develop and implement educational programs that train cardiovascular health care professionals to practice and facilitate mindfulness. Increasing the number of qualified practitioners is expected to contribute to the sustainable delivery and expansion of such programs in clinical and community settings. In addition, future efforts should focus on expanding the availability of mobile apps across different digital platforms and, ultimately, adapting and implementing the program in international settings. Such developments may facilitate broader accessibility and cross-cultural validation of mindfulness-based interventions for patients with chronic cardiopulmonary diseases.

Conclusions

This feasibility study suggests that an online mindfulness-based self-management program may help patients with pulmonary hypertension engage more effectively with symptoms, treatment-related side effects, and emotional distress. Participants reported preliminary improvements in perceived pain and aspects of quality of life, indicating potential psychological benefits. Further refinement of the program and evaluation in larger randomized controlled trials are needed to determine its efficacy and long-term impact.

Acknowledgments

We sincerely appreciate the English language editing support provided by Editage, which helped improve the clarity and readability of this manuscript.

Funding

This work was supported by JSPS KAKENHI grants 21H03243 and 23K21547.

Data Availability

The raw data analyzed qualitatively will not be disclosed to maintain the confidentiality of the responses, but the quantitatively analyzed datasets may be provided by the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: YT (lead), DF (equal), SP (equal)
Data curation: YT
Formal analysis: YT (lead), JM (equal)
Funding acquisition: YT
Investigation: YT (lead), JM (equal), AG (supporting), TI (supporting), HK (supporting)
Methodology: YT (lead), SP (equal), JM (supporting), DF (supporting)
Project administration: YT
Resources: YT (lead), SP (equal), AG (equal), JM (supporting), DF (supporting)
Supervision: DF (lead), SP (equal)
Validation: YT (lead), SP (equal), AG (equal), JM (supporting), DF (supporting)
Visualization: YT (lead), SP (supporting), JM (supporting), DF (supporting)
Writing – original draft: YT
Writing – review & editing: DF (lead), TK (supporting), MK (supporting)

Conflicts of Interest

TI receives lecture fees from MSD and Bayer outside the submitted work. YT receives lecture fees from MSD and Mochida Pharmaceutical and serves as an advisor for an international meeting organized by MSD, outside the submitted work. The other authors declare no conflicts.

References

1. Ogawa A, Ejiri K, Matsubara H. Long-term patient survival with idiopathic/heritable pulmonary arterial hypertension treated at a single center in Japan. Life Sci 2014 Nov 24;118(2):414-419. [doi: 10.1016/j.lfs.2014.01.077] [Medline: 24530872]

XSL•FO
RenderX

2. Humbert M, Lau EMT, Montani D, Jaïs X, Sitbon O, Simonneau G. Advances in therapeutic interventions for patients with pulmonary arterial hypertension. Circulation 2014 Dec 9;130(24):2189-2208. [doi: 10.1161/CIRCULATIONAHA.114.006974] [Medline: 25602947]
3. Aguirre-Camacho A, Moreno-Jiménez B. Depression and anxiety in patients with pulmonary hypertension: the role of life satisfaction and optimism. Psychosomatics 2018 Nov;59(6):575-583. [doi: 10.1016/j.psym.2018.04.002] [Medline: 29961653]
4. Funabashi S, Kataoka M, Inami T, et al. Depressive status in patients with chronic thromboembolic pulmonary hypertension. Circ J 2017 Jun 23;81(7):1051-1053. [doi: 10.1253/circj.CJ-17-0197] [Medline: 28529263]
5. Halimi L, Marin G, Molinari N, et al. Impact of psychological factors on the health-related quality of life of patients treated for pulmonary arterial hypertension. J Psychosom Res 2018 Feb;105:45-51. [doi: 10.1016/j.jpsychores.2017.12.003] [Medline: 29332633]
6. Harzheim D, Klose H, Pinado FP, et al. Anxiety and depression disorders in patients with pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension. Respir Res 2013 Oct 9;14(1):104. [doi: 10.1186/1465-9921-14-104] [Medline: 24107187]
7. Ivarsson B, Hesselstrand R, Rådegran G, Kjellström B. Health-related quality of life, treatment adherence and psychosocial support in patients with pulmonary arterial hypertension or chronic thromboembolic pulmonary hypertension. Chron Respir Dis 2019;16:1479972318787906. [doi: 10.1177/1479972318787906] [Medline: 30011997]
8. M M Vanhoof J, Delcroix M, Vandevelde E, et al. Emotional symptoms and quality of life in patients with pulmonary arterial hypertension. J Heart Lung Transplant 2014 Aug;33(8):800-808. [doi: 10.1016/j.healun.2014.04.003] [Medline: 24854567]
9. McCollister DH, Beutz M, McLaughlin V, et al. Depressive symptoms in pulmonary arterial hypertension: prevalence and association with functional status. Psychosomatics 2010;51(4):339-339. [doi: 10.1176/appi.psy.51.4.339] [Medline: 20587763]
10. Pfeuffer E, Krannich H, Halank M, et al. Anxiety, depression, and health-related QOL in patients diagnosed with PAH or CTEPH. Lung 2017 Dec;195(6):759-768. [doi: 10.1007/s00408-017-0052-z] [Medline: 28993877]
11. Somaini G, Hasler ED, Saxer S, et al. Prevalence of anxiety and depression in pulmonary hypertension and changes during therapy. Respiration 2016;91(5):359-366. [doi: 10.1159/000445805] [Medline: 27189373]
12. Tartavoulle TM. A predictive model of the effects of depression, anxiety, stress, 6-minute-walk distance, and social support on health-related quality of life in an adult pulmonary hypertension population. Clin Nurse Spec 2015;29(1):22-28. [doi: 10.1097/NUR.0000000000000099] [Medline: 25469437]
13. White J, Hopkins RO, Glissmeyer EW, Kitterman N, Elliott CG. Cognitive, emotional, and quality of life outcomes in patients with pulmonary arterial hypertension. Respir Res 2006 Mar 31;7(1):55. [doi: 10.1186/1465-9921-7-55] [Medline: 16579853]
14. Takita Y, Takeda Y, Fujisawa D, Kataoka M, Kawakami T, Doorenbos AZ. Depression, anxiety and psychological distress in patients with pulmonary hypertension: a mixed-methods study. BMJ Open Respir Res 2021 Apr;8(1):e000876. [doi: 10.1136/bmjresp-2021-000876] [Medline: 33926959]
15. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation for Everyday Life: Piatkus Books; 2004.
16. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness: Delacorte Press; 2013.
17. Segal ZV, Williams JMG, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression, 2nd edition: Guilford Press; 2013.
18. Lengacher CA, Kip KE, Barta M, et al. A pilot study evaluating the effect of mindfulness-based stress reduction on psychological status, physical status, salivary cortisol, and interleukin-6 among advanced-stage cancer patients and their caregivers. J Holist Nurs 2012 Sep;30(3):170-185. [doi: 10.1177/0898010111435949] [Medline: 22442202]
19. van den Hurk DGM, Schellekens MPJ, Molema J, Speckens AEM, van der Drift MA. Mindfulness-based stress reduction for lung cancer patients and their partners: results of a mixed methods pilot study. Palliat Med 2015 Jul;29(7):652-660. [doi: 10.1177/0269216315572720] [Medline: 25701663]
20. Johns SA, Beck-Coon K, Stutz PV, et al. Mindfulness training supports quality of life and advance care planning in adults with metastatic cancer and their caregivers: results of a pilot study. Am J Hosp Palliat Care 2020 Feb;37(2):88-99. [doi: 10.1177/1049909119862254] [Medline: 31378080]
21. Cottingham AH, Beck-Coon K, Bernat JK, et al. Addressing personal barriers to advance care planning: qualitative investigation of a mindfulness-based intervention for adults with cancer and their family caregivers. Palliat Support Care 2019 Jun;17(3):276-285. [doi: 10.1017/S1478951518000354] [Medline: 29880064]
22. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303. [doi: 10.2196/jmir.7487] [Medline: 28860106]
23. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253. [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]

XSL•FO
RenderX

24. Nasiri Z, Alavi M, Ghazavi Z, Rabiei K. The effectiveness of mindfulness-based intervention on perceived stress and perception of disease in patients with acute coronary syndrome. J Educ Health Promot 2020;9:130. [doi: 10.4103/jehp.jehp_660_19] [Medline: 32642486]
25. Abdul Manan H, Mir IA, Humayra S, Tee RY, Vasu DT. Effect of mindfulness-based interventions on anxiety, depression, and stress in patients with coronary artery disease: a systematic review and meta-analysis of randomized controlled trials. Front Psychol 2024;15:1435243. [doi: 10.3389/fpsyg.2024.1435243] [Medline: 39144586]
26. Cavalcante VN, Mesquita ET, Cavalcanti ACD, et al. Impact of a stress reduction, meditation, and mindfulness program in patients with chronic heart failure: a randomized controlled trial. Arq Bras Cardiol 2023 Oct;120(10):e20220768. [doi: 10.36660/abc.20220768] [Medline: 37909602]
27. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Heart J 2022 Oct 11;43(38):3618-3731. [doi: 10.1093/eurheartj/ehac237] [Medline: 36017548]
28. Ethical guidelines for life science and medical research involving human subjects. : Ministry of Education, Culture, Sports, Science and Technology; 2021 URL: https://www.mhlw.go.jp/content/001457376.pdf [accessed 2026-01-28]
29. Heo S, McSweeney J, Ounpraseuth S, Shaw-Devine A, Fier A, Moser DK. Testing a holistic meditation intervention to address psychosocial distress in patients with heart failure: a pilot study. J Cardiovasc Nurs 2018;33(2):126-134. [doi: 10.1097/JCN.0000000000000435] [Medline: 28661991]
30. Freedenberg VA, Thomas SA, Friedmann E. A pilot study of a mindfulness based stress reduction program in adolescents with implantable cardioverter defibrillators or pacemakers. Pediatr Cardiol 2015 Apr;36(4):786-795. [doi: 10.1007/s00246-014-1081-5] [Medline: 25519914]
31. Yorke J, Corris P, Gaine S, et al. emPHasis-10: development of a health-related quality of life measure in pulmonary hypertension. Eur Respir J 2014 Apr;43(4):1106-1113. [doi: 10.1183/09031936.00127113] [Medline: 24232702]
32. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
33. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
34. Connor KM, Davidson JRT. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
35. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
36. Krippendorff K. Content Analysis: An Introduction to Its Methodology, 3rd edition: SAGE Publications; 2018. [doi: 10.4135/9781071878781]
37. Boyle CC, Stanton AL, Ganz PA, Crespi CM, Bower JE. Improvements in emotion regulation following mindfulness meditation: effects on depressive symptoms and perceived stress in younger breast cancer survivors. J Consult Clin Psychol 2017 Apr;85(4):397-402. [doi: 10.1037/ccp0000186] [Medline: 28230391]
38. Hou Y, Zhao X, Lu M, Lei X, Wu Q, Wang X. Brief, one-on-one, telephone-adapted mindfulness-based stress reduction for patients undergoing percutaneous coronary intervention: a randomized controlled trial. Transl Behav Med 2019 Nov 25;9(6):1216-1223. [doi: 10.1093/tbm/ibz130] [Medline: 31504974]
39. Nakamura Y, Lipschitz DL, Kuhn R, Kinney AY, Donaldson GW. Investigating efficacy of two brief mind-body intervention programs for managing sleep disturbance in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv 2013 Jun;7(2):165-182. [doi: 10.1007/s11764-012-0252-8] [Medline: 23338490]
40. Nyklíček I, Dijksman SC, Lenders PJ, Fonteijn WA, Koolen JJ. A brief mindfulness based intervention for increase in emotional well-being and quality of life in percutaneous coronary intervention (PCI) patients: the MindfulHeart randomized controlled trial. J Behav Med 2014 Feb;37(1):135-144. [doi: 10.1007/s10865-012-9475-4] [Medline: 23180285]
41. Farb N, Daubenmier J, Price CJ, et al. Interoception, contemplative practice, and health. Front Psychol 2015;6:763. [doi: 10.3389/fpsyg.2015.00763] [Medline: 26106345]
42. Matzka M, Mayer H, Köck-Hódi S, et al. Relationship between resilience, psychological distress and physical activity in cancer patients: a cross-sectional observation study. PLOS ONE 2016;11(4):e0154496. [doi: 10.1371/journal.pone.0154496] [Medline: 27124466]
43. Koelmel E, Hughes AJ, Alschuler KN, Ehde DM. Resilience mediates the longitudinal relationships between social support and mental health outcomes in multiple sclerosis. Arch Phys Med Rehabil 2017 Jun;98(6):1139-1148. [doi: 10.1016/j.apmr.2016.09.127] [Medline: 27789238]

Abbreviations

CD-RISC: Connor-Davidson Resilience Scale
GAD-7: Generalized Anxiety Disorder 7-item scale
MBCT: mindfulness-based cognitive therapy
MBSR: mindfulness-based stress reduction
PAH: pulmonary arterial hypertension

XSL•FO
RenderX

PH: pulmonary hypertension
PHQ-9: Patient Health Questionnaire-9
QOL: quality of life

Edited by A Coristine; submitted 26.Jun.2025; peer-reviewed by AFM Giuliano, MB Torres-Rojas; accepted 16.Dec.2025; published 04.Feb.2026.

Please cite as:
Takita Y, Morishita J, Park S, Goda A, Inami T, Kikuchi H, Kohno T, Kataoka M, Fujisawa D
Mindfulness-Based Self-Management Program Using a Mobile App for Patients With Pulmonary Hypertension: Single-Arm Feasibility Study
JMIR Cardio 2026;10:e79639
URL: https://cardio.jmir.org/2026/1/e79639
doi:10.2196/79639

© Yuka Takita, Junko Morishita, Sunre Park, Ayumi Goda, Takumi Inami, Hanako Kikuchi, Takashi Kohno, Masaharu Kataoka, Daisuke Fujisawa. Originally published in JMIR Cardio (https://cardio.jmir.org), 4.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Barriers and Facilitators to Maintaining Physical Activity After an mHealth Intervention for People Poststroke or Transient Ischemic Attack: Reflexive Thematic Analysis

Hanna Lagerlund[1], MSc; David Moulaee Conradsson[1,2], PhD; Charlotte K Häger[3,4], PhD; Lydia Kwak[5], PhD; Lisa Holmlund[6], PhD

[1]Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Alfred Nobels Allé 23, 23100, Stockholm, Sweden

[2]Medical Unit Allied Health Professionals, Women's Health and Allied Health Professionals Theme, Karolinska University Hospital, Stockholm, Sweden

[3]Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Västerbotten, Sweden

[4]Department of Diagnostics and Intervention, Umeå University, Umeå, Västerbotten, Sweden

[5]Unit of Intervention and Implementation Research for Worker Health, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[6]Division of Occupational Therapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Hanna Lagerlund, MSc
Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Alfred Nobels Allé 23, 23100, Stockholm, Sweden

Abstract

Background: Regular physical activity is critical for preventing secondary stroke following a stroke or transient ischemic attack (TIA). Although mobile health (mHealth) interventions have shown promise for promoting short-term increases in physical activity, evidence on their long-term effects and the mechanisms that support sustained behavior change remains limited. In particular, little is known about how people poststroke or TIA integrate the skills, knowledge, and habits gained through mHealth interventions into their daily lives once structured intervention support ends.

Objective: This study aimed to explore the perceived barriers and facilitators of maintaining physical activity among individuals poststroke or TIA after completing an mHealth intervention.

Methods: A qualitative approach was used, involving a strategic sample of 12 participants recruited after they had completed a 6-month mHealth intervention for people poststroke or TIA. The intervention included supervised physical therapy with mHealth support for physical exercises and behavior change (eg, counseling, goal-setting, and self-monitoring), followed by a 6-month postintervention period with access to self-managed mHealth support. To enable richness and depth in participants' accounts, the dataset consisted of 2 semistructured interviews conducted 3 months and 6 months after completing the intervention, along with participant-generated photographs. Between the interviews, participants took photos to reflect their experiences of maintaining physical activity after the intervention. These images served as prompts for dialog and reflection during the second interview. Data were analyzed using reflexive thematic analysis.

Results: We generated 3 themes: building experience and knowledge to maintain physical activity, staying physically active in a complex life situation, and the meaning of context for maintaining physical activity after the intervention. Barriers and facilitators were conceptually integrated into a comprehensive understanding of maintaining physical activity, symbolized by the metaphor of a soap bubble, which requires persistence and a supportive environment to stay afloat. The complexity of participants' health and life situations created barriers to maintaining physical activity. To overcome these barriers, developing personal strategies within a supportive social and physical context was crucial. This development was facilitated by the mHealth intervention, which enabled knowledge acquisition of the principles of physical activity after stroke or TIA, along with increased awareness of its health benefits.

Conclusions: mHealth interventions for people poststroke and TIA can serve as a catalyst for physical activity engagement and an enhancer of the knowledge and experience necessary to maintain physical activity after the intervention. Despite health-related and contextual barriers, participants may use personalized strategies, supported by their social and environmental context, to navigate these challenges. These insights highlight opportunities for future mHealth interventions to strengthen the interaction between individuals and their environment and empower tailored strategies for behavior change and long-term physical activity maintenance.

Trial Registration: ClinicalTrials.gov NCT05111951; https://clinicaltrials.gov/study/NCT05111951

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.1186/s12883-023-03163-0

(JMIR Cardio 2026;10:e87493) doi:10.2196/87493

KEYWORDS

digital interventions; digital health; mobile health; mobile apps; photo-elicitation; physical activity maintenance; stroke

Introduction

Stroke is the third leading cause of death and disability globally, representing a major public health challenge that demands effective preventive strategies [1]. As a modifiable risk factor for recurrent stroke and transient ischemic attack (TIA), physical activity plays a vital role in secondary stroke prevention [2]. However, many individuals poststroke and TIA fail to meet recommended physical activity levels [3,4]. This highlights the need for interventions that not only promote physical activity but also support its long-term maintenance [5,6].

Physical activity refers to all movements produced by skeletal muscles that require energy expenditure, including leisure activities, commuting, and household chores [7,8]. In contrast, physical exercise is a subcategory of physical activity that is structured, goal-directed, and aimed at enhancing physical fitness [7,8]. In the context of digital transformation of care, mobile health (mHealth) has emerged as a promising approach to promoting physical activity among people poststroke or TIA [9]. There is recent evidence that mHealth interventions can support people poststroke or TIA to establish exercise routines and increase physical activity in the short term [10,11], whereas strategies to sustain these improvements over time remain underexplored [5]. Furthermore, most mHealth trials rely on self-management alone [11], which contrasts with the expressed need for therapeutic support from professionals experienced in stroke rehabilitation, as reported by people poststroke or TIA [12]. In line with this, a recent overview showed that multicomponent interventions combining therapist-led sessions with self-management provide moderate evidence for improving physical activity and reducing cardiac events poststroke or TIA [13]. Therefore, to develop mHealth interventions that effectively support long-term physical activity after stroke or TIA, it is crucial to deepen our understanding of how people poststroke or TIA experience the maintenance of physical activity through self-management after completing an intervention that included therapeutic support. Moreover, more knowledge is needed about how people poststroke or TIA integrate the skills, knowledge, and habits acquired through mHealth interventions into their daily lives once structured support ends.

Physical activity maintenance could be operationalized using behavioral indicators such as meeting predefined thresholds (eg, 150 minutes of physical activity per week), consistency over a specified observation period, and regularity of engagement [14]. Several factors are known to influence physical activity engagement among individuals poststroke and TIA. Facilitators include peer support and self-efficacy for participation in physical activity, while barriers include stroke-related impairments and inadequate public transportation to physical activity venues [15,16]. Maintaining physical activity over time may involve distinct factors to those influencing the initiation of physical activity [17]. Developing a self-identity as a physically active person and incorporating flexibility into routines have, for example, been identified as facilitators specific to long-term physical activity maintenance in adults [17]. Moreover, among older adults, scheduled physical activity, encouragement from spouses, and access to both indoor and outdoor environments have been shown to facilitate the maintenance of physical activity, while the absence of these factors acted as barriers [18]. These findings of person-environment interaction in maintaining physical activity are in line with the social cognitive theory (SCT), which posits that environmental support and individual self-regulation are important for behavior change [19]. SCT also emphasizes how personal factors, a person's behavior (what they do or think), and the environment in which they act interact reciprocally, meaning that changes in one of these factors influence the others [19]. To the best of our knowledge, there is currently no research that specifically addresses the maintenance of physical activity following an mHealth intervention in individuals poststroke or TIA.

This study builds on a telehealth intervention from Australia, *i-REBOUND – Let's Get Moving!*, designed to support the promotion and maintenance of physical activity in people poststroke or TIA through supervised exercise and support for self-management through behavior change techniques [20]. In prior work by our group, i-REBOUND was adapted into a digital intervention combining physical therapist–led sessions (supervised exercise and counseling) and self-managed mHealth support (eg, activity diary, prerecorded exercise videos), delivered via a mobile phone app [12]. The mHealth version of the i-REBOUND intervention was found to be feasible, acceptable, and successful at delivering the intervention in an mHealth format [21,22]. To further refine the mHealth version of the i-REBOUND intervention and address the knowledge gap regarding the maintenance of physical activity following mHealth interventions, it is important to understand which factors positively and negatively influence maintained physical activity. This study aimed to explore the perceived barriers and facilitators to maintaining physical activity among individuals poststroke or TIA after completing the mHealth version of the i-REBOUND intervention.

Methods

Study Design

This study is grounded in a constructivist epistemology, in which knowledge is understood as coconstructed through social interaction between participants and researchers rather than discovered as an objective truth. The dataset was generated through repeated semistructured interviews and

XSL•FO
RenderX

participant-generated photos [23], and a reflexive thematic approach was applied in the data analysis [24,25].

The authors form an interdisciplinary team of 4 women and 1 man, all based in academic settings at Swedish universities. Four of the authors also have experience in clinical practice within physiotherapy (CKH, DMC, and HL) or occupational therapy (LH). All are experienced in intervention and implementation research, including mHealth (CKH and DMC) and knowledge of behavior change theories (HL and LK). In line with constructivism and reflexive thematic analysis [26], themes are viewed as actively generated through an interpretative and reflexive analytic process that draws on participants' accounts and researchers' positionality. Participants with experience of the mHealth version of the i-REBOUND intervention were integral to the knowledge production process, and we acknowledged our active role as researchers in shaping the analysis. The study's reporting follows the Reflexive Thematic Analysis Reporting Guidelines [27].

Participants

We strategically recruited 12 participants (10 women, 2 men) from the experimental group (n=57) of a feasibility randomized controlled trial (n=114) after participating in the mHealth version of the i-REBOUND intervention for people poststroke and TIA [21]. Inclusion criteria for the feasibility randomized controlled trial were adults who (1) experienced a stroke or TIA between 3 months and 10 years prior to the study, (2) lived at home, (3) had access to a reliable internet connection, (4) could walk short distances indoors with or without a walking device, (5) could use a smartphone, and (6) could self-identify through a Swedish secure self-authentication app [28]. Exclusion criteria were (1) currently meeting recommended physical activity levels (ie, at least 150 minutes per week of moderate physical activity or at least 75 minutes per week of vigorous intensity physical activity [29]) or had (2) severe health conditions and/or (3) significant cognitive impairments [28]. Physical activity levels of the participants were assessed via a telephone interview, and individuals were eligible for inclusion if they self-reported not meeting the recommended 150 minutes of moderate-intensity physical activity per week based on frequency, intensity, and duration. The strategic sampling aimed to achieve diversity in age, sex, physical activity level, geography, and stroke or TIA diagnosis. Twelve participants were identified and agreed to participate. Among the recruited participants, 10 regions across Sweden were represented, and the participants were aged 56 years through 80 years at the start of the intervention. Participant characteristics are presented in Table 1.

XSL•FO
RenderX

Table . Participant characteristics at baseline (n=12), recruited from the experimental group of a feasibility randomized controlled trial.

Characteristics	Value
Age[a] (years), mean (min-max)	70 (56 - 80)
Women, n (%)	10 (83)
Clinical diagnosis, n (%)	
Stroke	8 (67)
TIA[b]	4 (33)
Duration since stroke/TIA diagnosis (years), mean (min-max)	4 (0.67 - 9.67)
Having a comorbidity[c] (yes), n (%)	11 (92)
Highest level of education, n (%)	
Primary school	1 (8)
Secondary school	6 (50)
University	5 (42)
Employed (yes), n (%)	4 (33)
Living with someone (yes), n (%)	8 (67)
Living in a rural area (yes), n (%)	3 (25)
Modified Rankin Scale[d] (0 - 6), mean (min-max)	1 (0 - 1)
Fatigue Severity Scale[e] (1-7), mean (min-max)	3.7 (1.8 - 5.4)
Previous experience with digital health care services, n (%)	7 (58)
Daily use of internet and mobile apps, n (%)	8 (67)
Steps per day, mean (min-max)	6901 (1048 - 13,897)

[a]All demographic data were collected at baseline, between September and November 2022, prior to the start of the mobile health version of the i-REBOUND intervention.

[b]TIA: transient ischemic attack.

[c]Comorbidity is defined as the co-existence of one or more conditions alongside the primary condition (stroke/TIA) [30].

[d]Assessment of the level of stroke disability according to the Modified Rankin Scale [31].

[e]Fatigue was assessed using the 9-item version of Fatigue Severity Scale [32].

The mHealth Version of the i-REBOUND Intervention

For full intervention details, see the study protocol [28]. The mHealth version of the i-REBOUND intervention aimed to support home-based exercise and physical activity for individuals poststroke or TIA. The intervention was delivered through an mHealth app called the STAAR (Stroke Treatment through Active and Accessible Rehabilitation) app. During the 6-month intervention, participants received digital supervised support from a physical therapist and had access to self-managed mHealth support. The participants attended weekly exercise sessions of moderate to vigorous intensity (months 1 to 3: two 25-minutes sessions per week; months 4 to 6: one 25-minute session per week) and monthly individual 30-minute counseling sessions led by a physical therapist. The counseling sessions focused on setting physical activity goals, discussing progression, and developing self-management strategies to support maintained physical activity engagement. Examples of goals were “Gym training twice a week on Mondays and Wednesdays. 20 minutes of aerobic exercise and 40 minutes of strength training. Borg scale 13 - 14.” and “Brisk walks two to three times per week, 30 to 60 minutes per session.” To further support the promotion of physical activity, the app provided self-managed mHealth support, including educational videos, an activity diary for self-monitoring, and prerecorded exercise videos. During the intervention, a chat function enabled contact with the physical therapist. An action plan, including long-term goals, for maintaining engagement in physical activity postintervention was formulated with support of the physical therapist at the final counseling session. After the intervention, participants continued to have access to self-managed mHealth support, although no further contact with the physical therapist was provided.

Dataset Generation

Semistructured interviews were conducted in Swedish at 3 months and 6 months postintervention. Interviews were carried out via digital video calls by the first author, a PhD candidate and experienced geriatric physical therapist, between June 2023 and November 2023. In the 6-month interview, participant-generated photo elicitation was used [23]. The first author had brief contact with the participants prior to the interviews, due to her involvement in the data collection in the feasibility randomized controlled trial.

XSL•FO
RenderX

The first and last authors developed and pilot-tested the interview guides. The 3-month interview covered the following conceptual areas: (1) the participants' perceptions of maintaining physical activity after the intervention, (2) barriers and facilitators to maintaining physical activity after the intervention, and (3) expectations and concerns about maintaining physical activity after the intervention. Examples of interview questions were "What has been your experience of being physically active after the intervention ended?" and "Can you tell me what facilitated regular physical activity for you after the intervention ended?" Before the 6-month interview, the participants were encouraged to take 3 to 5 photos reflecting how they perceived maintaining their physical activity after completing the mHealth version of the i-REBOUND intervention. After a brief reflection on their current physical activity level, the participants were asked to reflect on one photo at a time. Open-ended questions were inspired by Green et al [33] and included "Tell me about this photo—why did you choose to bring this photo?" "Tell me what could prevent you from doing this activity regularly?" and "Was there anything missing in the intervention to make this activity regular?" To enhance the richness of the descriptions, participants were prompted with follow-up questions and encouraged to elaborate on their responses using examples of situations. The interviews lasted between 40 minutes and 90 minutes and were audio-recorded. Field notes were written after each interview. Interviews were transcribed and anonymized before the data analysis.

Data Analysis

Analysis followed reflexive thematic analysis, based on the 6-phase process outlined by Braun and Clark [24,25]. The primary focus was on the transcribed text; however, photos were analyzed alongside participants' interviews to add context and generate a richer understanding of the participants' verbal accounts [34]. The software program N-Vivo 15 was used to organize the transcribed data. In phase one, *familiarization with data*, the first author familiarized herself with the dataset by carefully reading and listening to all interviews while simultaneously taking field notes. In this phase, discussions were continuously held with the last author, who read a sample of the interviews. In phase two, *generating initial codes*, the first author, in collaboration with the last author, identified and labeled parts of the text that captured meaningful experiences related to the research aim. The third phase, *generating initial themes*, included reflexive discussions between the first and the last authors in which the initial codes were reviewed and refined through merging, renaming, and grouping codes to explore patterns of meaning across the dataset. Phases four and five, *theme development and review* and *theme refinement*, were carried out through a continuous iterative process with ongoing discussions among all authors, involving data, codes, and themes. The focus was on identifying clear patterns within each theme, ensuring conceptual consistency and coherence, and naming the themes [26]. In these later phases, photos were used to provide an overall interpretation of physical activity maintenance and to develop themes. For example, a reflexive discussion around a photo of a soap bubble and the story aligned with it helped capture an overall understanding of the complexity of maintaining physical activity and the environmental influence on this process. Further, there were many photos illustrating aspects of the environment. Through reflexive discussions about what these photos and stories meant for staying physically active after an mHealth intervention, we could further develop themes. In analyzing person-environment interactions, an existing behavior change theory, SCT [19], was used to inform data interpretation and further develop themes in line with reflexive thematic analysis [26,27]. For example, photos and stories about social and physical environments were useful for understanding aspects such as reinforcement and the role of social support in the first and third themes. In this way, the photos not only contextualized participants' narratives but also contributed analytically by expanding the understanding of barriers and facilitators to maintaining physical activity after the intervention. In phase six, *writing up*, the final report was produced, and all authors read, discussed, and contributed to the writing through several rounds of the manuscript.

Ethical Considerations

Ethical approval for the study was obtained from the Swedish Ethical Review Authority (dnr 2020‐05062 and 2022-07270-02). The research followed the Declaration of Helsinki, and careful consideration was given to ensure the integrity and well-being of the participants throughout the research process. Prior to their consent, all participants were provided with verbal and written information about the study, including their right to withdraw their participation at any time. No compensation for participation was provided. After transcription, the data were de-identified by removing identifiable information from the transcripts, and photographs were selected to ensure participants' confidentiality.

Results

Overview

The overall understanding of maintaining physical activity is symbolized by a photo of a soap bubble, first illustrated by a participant emphasizing the inevitable fragility of staying physically active (Figure 1): "For me this soap bubble is the a...almost impossible, a soap bubble doesn't last forever, I mean it really doesn't. But for how long can you hold on, for how long can you actually hold on" (female participant, 67 years old, TIA). Intervention mechanisms and personal factors contributed to the persistence of the soap film. At the same time, contextual conditions created an environment that kept the bubble afloat, even in the face of barriers in participants' life situations. We generated 3 themes through analysis, illustrating the barriers and facilitators to maintaining physical activity after completing the i-REBOUND intervention (Table 2).

XSL•FO
RenderX

Figure 1. A soap bubble illustrating the inevitable fragility of maintaining physical activity (Female participant, 67 years old, transient ischemic attack).



XSL•FO
RenderX

Table . Themes, subthemes, and barriers and facilitators to maintaining physical activity after completing the mobile health (mHealth) version of the i-REBOUND intervention.

Themes and subthemes	Barriers	Facilitators
Building experience and knowledge to maintain physical activity	• Limited experience with physical exercise	• Regular experience with physical exercise over the 6-month intervention • Positive experience with the health benefits of physical activity • Knowledge acquisition of principles of physical exercise after stroke and TIA[a] • Knowledge acquisition on how to use daily routines and activities as a resource of physical activity • Previous experiences with physical exercise after stroke or TIA
Staying physically active in a complex life situation		
Integrating physical activity into daily life amid health challenges	• Health impact of poststroke impairments and comorbidity • Impact of engagements, roles, and responsibilities in daily life • Lack of personal strategies for maintaining physical activity	—[b]
Strategies for integrating physical activity into daily life	• Minimal added value of the self-managed support in the STAAR[c] app	• Belief in one's capability to be physically active • Management of poststroke impairments and comorbidities • Self-monitoring through hand-written exercise logs or the step counter on the phone • Adhering to a structured yet feasible action plan
The meaning of context for maintaining physical activity after the intervention	• Discontinued support from a physical therapist • Weather and seasonal changes	• Involvement of significant others in the social network as physical activity companions • A supportive social context encouraging engagement in physical activity • A designated place for performing physical activity

[a]TIA: transient ischemic attack.

[b]In this subtheme, only barriers were described.

[c]STAAR: Stroke Treatment through Active and Accessible Rehabilitation.

Building Experience and Knowledge to Maintain Physical Activity

A persistent soap film served as a metaphor for gaining experience and acquiring knowledge through participation in the intervention. The continuous interaction with the physical therapist during individual counseling and regular exercise sessions over 6 months provided opportunities for habit formation. The regular engagement in exercise also enabled the participants to experience the benefits of physical activity for health and managing daily life. Along with a deeper understanding of the principles of physical exercise, the intervention components and mechanisms of change supported a stable foundation to maintain regular physical activity after the intervention.

The regular physical exercise sessions at higher intensity levels had served as a catalyst for physical activity and reinforced habit formation for continued activity after the intervention. In addition, being regularly active over time enabled the participants to experience and recognize the health benefits of physical activity. Examples of such benefits included reduced pain levels, increased energy for daily activities, and an overall feeling of improved well-being. The catalyst for change through these experiences was felt to be beneficial regardless of previous engagement in physical activity. One participant shared how the intervention was an awakening of the benefits of being regularly active: "I mean, I did move around before as well, but this has still given me a bit of…a little step forward, in some way. I think so. Uh, a wake-up call. That this is…this is good. It's good for me to do these things" (female participant, 76 years old, stroke). In line with experiencing health benefits, the participants also became aware of the importance of increased strength for managing daily activities. This awareness was reflected in a photo of stairs (Figure 2). The participant lived in a building without an elevator, and the photo symbolized how regular physical activity was a prerequisite to sustaining the ability to climb stairs and accessing the apartment. This

XSL•FO
RenderX

example illustrated how maintaining physical activity was perceived not only as beneficial for health but also as a means for sustaining important abilities for their daily life.

During the intervention, the participants developed an understanding of the importance of exercise intensity and the benefits of having an increased heart rate during physical exercise. Participants used this knowledge to adapt their physical exercise routines to meet recommendations for physical activity. Moreover, participants learned how to use daily activities as opportunities for physical activity. Walking, in particular, was an important and easily accessible activity that could be adapted, for example, by replacing slow-paced dog walks with brisk power walks without the dog. Another example of modifying an existing routine, illustrated by a photo of a swimming pool (Figure 3), was the incorporation of faster swimming intervals to enhance intensity. In addition, the introduction of tools, such as the Borg scale [35], and specific exercises provided a foundation for a personalized exercise routine and set the stage for progression. The familiarity and simplicity of the intervention's exercises supported maintenance, as they were easy to remember and carry out, and the Borg scale was a practical tool for the participants to monitor and achieve their target intensity levels.

XSL•FO
RenderX

Figure 2. A staircase reinforcing physical activity maintenance for sustained ability to climb stairs (female participant, 66 years old, stroke).



The participants entered the intervention with different levels of prior knowledge for their knowledge acquisition. Participants with previous exercise experience could continue to develop knowledge, and the intervention's supervised support reinforced their understanding. In addition, the supervised sessions helped these participants develop realistic expectations for physical exercise, adapted to their age and current physical ability. One participant reflected: "This thing about setting the training at a,

XSL•FO
RenderX

yeah…a level that's just right. That made it possible to do things differently. It just felt…I mean, mentally it was different" (male participant, 70 years old, TIA). As the quote illustrates, the knowledge acquired during the intervention prompted a shift in mindset that facilitated acceptance of exercise at a more manageable level. For the participants with limited experience of physical exercise, the intervention served as an introduction to establishing exercise routines and contributed to the acquisition of knowledge on exercise and physical activity, which supported their ability to maintain physical activity after the intervention.

XSL•FO
RenderX

Figure 3. A swimming pool illustrating a modified exercise routine for increased exercise intensity (female participant, 76 years old, stroke).



XSL•FO
RenderX

Staying Physically Active in a Complex Life Situation

Overview

Maintaining physical activity after the intervention could be challenging due to participants' health and contextual circumstances. The participants' life situations often involved poststroke impairments, comorbid conditions, and daily roles and responsibilities that continuously competed with time and energy for physical activity. Facilitators to overcome these barriers, some of them introduced by the intervention, included a variety of personal strategies to support maintained physical activity. These barriers and facilitators are illustrated in two subthemes: *Integrating physical activity into daily life amid health challenges* and *Strategies for integrating physical activity into daily life*.

Integrating Physical Activity Into Daily Life Amid Health Challenges

Health challenges posed considerable barriers to maintaining physical activity after the intervention. Comorbidities, such as heart disease, arthritis, and cancer, meant a balancing act between how the consequences of their condition affected physical ability and the struggle to maintain being physically active. One participant described how multiple hospital visits, medical treatment, and an exercise ban prescribed by the physician disrupted all plans for exercising: "Because of all that stuff—I was about to say—with atrial fibrillation and all that…and the inflammation I had, which threw off…or ruined all that with the training. It [medical problems] has continued…" (male participant, 70 years old, TIA). In addition to disrupting physical activity, health challenges required the participants to engage in physical activity at a lower level than originally intended, according to the goals set at the end of the intervention. Fatigue and short-term illness, such as colds, infections, and musculoskeletal pain, forced the participants to rebuild their physical activity routines repeatedly, as one participant said: "I've been interrupted all the time, that's the thing. As soon as I start and get going a little then [laughs]…something happens" (female participant, 71 years old, stroke). These constant interruptions both affected the ability to maintain physical activity and, at times, undermined the motivation to continue the efforts.

Maintaining physical activity was also influenced by a pull of other engagements, roles, and responsibilities of daily life, which could function as both barriers and facilitators for physical activity. On the one hand, engaging in activities such as gardening and caring for grandchildren facilitated physical activity. One participant shared a photo of 2 pairs of shoes placed next to each other (Figure 4), illustrating how activities with her grandchildren, such as going for a bike ride with them instead of driving, facilitated physical activity. However, for maintaining intensive exercise, it was necessary to integrate such exercise sessions into daily life at large. On the other hand, participants also reflected that the pull of sedentary daily activities was a barrier, sometimes stronger than the push to engage in regular physical activity. This insight was coupled with a sense of failure and personal responsibility for not meeting the expectations after the intervention. One participant exemplified this struggle with a photo of her sofa (Figure 5):

> *I don't mind engaging in physical exercise, but it is…It's so hard to make it happen at home, I don't know why, but I'm kind of generally lazy and sluggish, I have to say, so it ends up with me sitting on the couch watching TV instead. Unfortunately.* [Female participant, 80 years old, TIA]

XSL•FO
RenderX

Figure 4. Shoes side by side illustrating engagement with grandchildren as a facilitator of physical activity (female participant, 67 years old, transient ischemic attack).



XSL•FO
RenderX

Figure 5. A living room sofa symbolizing sedentary routines competing with physical activity engagement (female participant, 80 years old, transient ischemic attack).



Similarly, the participants' roles and responsibilities could pose barriers for maintaining regular physical activity; for example, one woman with the responsibility of caring for a sick spouse said: "Yeah, … taking care of the home. Mhm. I have a sick husband, so…Everything kind of falls on me, in a way" (female participant, 76 years old, stroke). The responsibilities for household chores meant that the routines established during the intervention were put on hold for her, and given her husband's condition, leaving the house was difficult. Overall, the stories illustrate how maintaining physical activity was negotiated daily in terms of participants' time, energy, and other priorities. The struggle to allocate time for physical exercise was illustrated in a photo of 2 towels (Figure 6). The participant explained that an exercise session, including showering, decreased the time available for other activities she had to do. A lack of personal strategies to navigate through these difficulties, including health challenges, was a barrier to maintaining physical activity.

XSL•FO
RenderX

Figure 6. Two towels on a towel holder symbolizing the struggle to allocate time for physical exercise (female participant, 76 years old, transient ischemic attack).



XSL•FO
RenderX

Strategies for Integrating Physical Activity Into Daily Life

In contrast to the stories about participants' challenges with maintaining physical activity, participants shared strategies developed to support their regular engagement in physical activity after the intervention. The goals established at the end of the intervention provided each participant with an individualized action plan for ongoing exercise and physical activity. To stay aligned with these goals, one participant shared a visual strategy of integrating them into her daily routine: "I've got them here all the time. Like…I've written them down, so I have them…together with Borg, I have them on my desk there. All the time. So I look at them [when she passes by]" (female participant, 75 years old, stroke). The attitude expressed in the quote reflects a belief in her capability to use the tools provided by the intervention and daily visual cues, both of which facilitated maintaining physical activity after the intervention. Another participant shared a sense of control over maintaining physical activity after a short-term illness: "It's like always, when you get going with something like this, you get sick or something. But I feel, it's fine, you know. I just continue from where I was*"* (female participant, 56 years old, stroke). In cases where participants struggled with long-term physical impairments and fatigue after stroke or TIA, a similar capability to develop strategies to overcome challenges was identified, including taking medication before exercise, balancing activity and rest, or prioritizing activities in daily life, despite their fatigue.

Although some strategies were developed by the participants themselves, other strategies had been initiated during the intervention. One such strategy was self-monitoring through the activity diary in the STAAR app. However, participants saw reporting in the app as an extra task with little added value compared with other self-monitoring methods, and it was therefore not considered a facilitator for maintaining physical activity. Instead, facilitators were tools such as handwritten exercise logs or step counters on their phones. These easily accessible tools provided both reinforcement and validation of what had been achieved while also acting as a motivational prompt, indicating the need to engage in more physical activity in order to reach the week's goal. Similarly, the prerecorded exercise videos in the STAAR app were used only to a limited extent, as they did not meet the participants' needs for intensity and variation. Consequently, other digital resources were explored, such as live exercise sessions broadcasted by the public service morning show. Both the activity diary for self-monitoring and the exercise videos in the STAAR app shared a common challenge: They were underused, and participants developed alternative strategies that supported them to maintain physical activity after the intervention. Among these strategies, participants emphasized the importance of planning and routine, such as following a preplanned weekly schedule or exercising in the mornings to ensure that physical activity happened. One participant shared adhering to a structured yet feasible action plan:

> *So, if I've decided on Tuesday and Thursday, and Tuesday turns out crap, then I just do Wednesday instead. It's not the end of the world. I mean…just switch the day. You don't have to be that rigid. Just because it didn't work on Tuesday doesn't mean I have to wait until next Tuesday. I mean…I think that's exactly why you have these planned routines—so it's actually feasible.* [Female participant, 56 years old, stroke]

As illustrated by the quote, adapting the action plan for exercise enabled the participants to overcome barriers like fatigue or difficulties balancing physical activities with other engagement, roles, and responsibilities in daily life.

The Meaning of Context for Maintaining Physical Activity After the Intervention

The participants' context—including dynamic interactions with their social and physical environments—was crucial for keeping the physical activity bubble afloat after the intervention. Within their immediate context, they identified significant others who could ease the transition from the physical therapist's supportive role, as well as meaningful places that offered favorable conditions to maintain physical activity.

Being part of the intervention entailed continuous social support for physical activity over 6 months. At completion, the participants missed both the guidance from the physical therapist and the social context provided by the regular exercise sessions. However, the participants identified alternative resources within their social context that facilitated the maintenance of physical activity after the intervention. One participant shared how taking part in the intervention made him realize the importance of having a companion to make it to the gym and identified a key person within his social network. He said:

> *My wife and I go to the gym two or three times a week, and it's important that she's with me. Otherwise, I might not go. She's great! I was thinking maybe I needed a personal trainer, but I don't. I just need someone to go with. And now I have that.* [Male participant, 56 years old, stroke]

In this way, having scheduled sessions with a companion fostered a sense of accountability, as someone was expecting the participant to attend. In addition, engaging in physical exercise with a companion provided a person who checked in on progress, contributed to a sense of safety, and gave feedback—similar to the support offered by the physical therapist during the intervention. Exercising with another person also brought joy and added value beyond physical activity itself. One participant described how the exercise class helped her build relationships and make friends, which facilitated her continued participation. Moreover, encouragement from significant others, such as friends, neighbors, and family members, facilitated maintaining physical activity by reinforcing the participants' decision to stay active and strengthening their sense of doing something positive for themselves. One participant illustrated this motivational support by quoting how her children encouraged her to prioritize engagement in physical exercise:

> *"'No, don't skip your physical exercise, we can meet up later' or 'No, do that first, it's what matters most to you' and stuff like that. They always cheer me on.*

XSL•FO
RenderX

> *They get that this has been really good for me, and they see that I've become a…I think I've changed quite a bit. Not in huge ways, but yeah, they can see that physical exercise really does me good.* [Female participant, 67 years old, TIA]

Alongside the social context, having a designated place to engage in physical activity served as a valuable resource that facilitated maintenance following the intervention. One factor for this place was accessibility, including geographical proximity and ease of access via transportation, contributing to a lowered threshold for exercising, as illustrated by one participant: "I like it when it's nearby and easy to get to and all that. That's kind of my problem—when it gets complicated or takes longer than I had in mind, I lose motivation" (female participant, 67 years old, TIA). The outdoor environment was one such example. In addition, engaging in physical activity outdoors provided mental rest through its calm and quiet atmosphere—something particularly valued by participants experiencing fatigue. It could also provide a sense of familiarity; one participant described feelings of comfort in her outdoor walking routine, as she was well-acquainted with the route's length and the time it typically took to complete. In one photo of a gravel trail (Figure 7), a participant captured how outside places could offer a meaningful nature experience and elaborated on how seasonal changes and the beautiful wildlife inspired her to take daily walks. However, using outdoor environments for daily physical activity was also associated with barriers related to seasonal and weather changes, especially rainy days or when the roads were slippery in the winter. For the participants, these barriers led to disruptions in their exercise routines—sometimes resulting in doing less than planned and, at other times, failing to adhere to the goals they had set at the end of the intervention. To overcome the weather and seasonal challenges, the participants needed to adapt their physical activity behaviors. One participant shared a photo of a kicksled (Figure 8) that exemplified this adapted behavior and how a kicksled enabled outdoor physical activity during the winter. For others, a pair of stable walking shoes or walking sticks helped maintain physical activity, through providing safety and inspiring them to go out for a walk regardless of the weather.

XSL•FO
RenderX

Figure 7. A gravel trail illustrating a designated place facilitating a meaningful physical activity experience (female participant, 75 years old, stroke).



XSL•FO
RenderX

Figure 8. A kicksled illustrating an adaptation to maintain physical activity during the winter season (female participant, 76 years old, stroke).



XSL•FO
RenderX

Discussion

Principal Findings

This study provides insights into barriers and facilitators of maintaining physical activity after completing an mHealth intervention from the perspective of individuals poststroke or TIA. The findings show how maintaining physical activity following an mHealth intervention is founded on a reciprocal relationship among the personal factors, behavior, and the environment. These findings align with behavioral change theories such as SCT [19], and the discussion will focus on aspects of personal, environmental, and behavioral factors for maintaining physical activity after an mHealth intervention.

This study found that personal factors, including positive experiences with regular physical exercise and knowledge acquired through interaction with a physical therapist, supported awareness of the health benefits of physical activity and habit formation. Consequently, these factors helped individuals incorporate physical activity into their daily routines and adjust their behaviors to follow physical activity guidelines. Similar findings were reported by Wahlich et al [36] who found that inactive mid-life and older adults made lifestyle changes in their daily lives and increased their awareness of the importance of physical activity after a pedometer-based walking intervention. Although confirming the findings of Wahlich et al [36], our study also demonstrated that participants not only acquired knowledge but also actively applied it by developing and adapting individual exercise routines after the intervention. This process was supported during the intervention by the creation of a social environment, which included support provided by the physical therapist who had extensive experience in stroke rehabilitation and played a key role by providing knowledge, reinforcement, and encouragement to maintain being physically active. After the intervention, this external support provided by the social environment was succeeded by participants' own experiences with positive outcomes, such as reduced pain and improved well-being, which, in line with SCT [37], further reinforced them to maintain regular physical activity.

The discontinuation of support from the physical therapist was perceived as a barrier to continued engagement in physical activity after the intervention. Supervision is recognized as an important component to supporting behavior change in individuals poststroke [38] but is inherently time-limited and resource-intensive. These constraints highlights the importance of identifying and strengthening alternative support once structured supervision ends. Our findings show how participants identified resources within their social network to substitute the physical therapist's supportive role during the intervention. These results are in line with previous findings highlighting that encouragement from significant others (eg, neighbors, friends, and family), as well as engaging in physical activity with a companion can facilitate maintaining physical activity [17,18,36]. The transition of support to the individual's own social network suggests, consistent with SCT [39], that individuals poststroke or TIA can sustain behavior change by developing alternative sources of social support within their immediate context following an mHealth intervention. This is an important finding, highlighting the need to incorporate strategies and training into mHealth interventions that help individuals integrate physical activity into daily life, supported by their social and environmental contexts.

Participants developed and used various personal strategies, some introduced during the intervention, to facilitate maintenance of physical activity. One such strategy was the self-monitoring of physical activity, where participants tracked their activity levels. The reported preference for using step counters on their mobile phone to monitor their physical activity was consistent with findings on strategies that facilitated physical activity maintenance in adult and older populations [17,18]. Handwritten exercise logs were preferably used, while the self-monitoring tool within the app was used to a lesser extent, possibly because it did not fully meet the needs of people poststroke or TIA. Our findings align with a previous quantitative evaluation of the mHealth version of the i-REBOUND intervention, which showed a decline in engagement with the app's self-managed tools over time [40]. This decline may reflect limited integration of key features known to enhance app engagement, such as personalization, adaptive content, and continuous feedback [41]. Although this study provides new insights into personal and environmental factors influencing maintenance—particularly those not directly related to the digital tools—further research is needed to better support long-term engagement in physical activity through mHealth solutions after stroke or TIA. Future refinements of the app may benefit from the development of more personalized content with regular updates and from the integration of automated self-monitoring of daily step counts or minutes in moderate physical activity, thereby improving its usability and relevance for this population. In SCT, self-monitoring is an important self-regulatory strategy that enables individuals to track their behavior, evaluate progress, and make necessary adjustments to achieve their goals [42]. In addition, SCT also emphasizes that maintaining physical activity requires the development of other self-regulatory capabilities, such as goal-setting and planning, as well to help individuals navigate barriers and maintain motivation [39]. These self-regulatory strategies were reflected in this study, demonstrating how participants applied the knowledge acquired from the intervention, along with their prior experience, and developed strategies for overcoming barriers. However, this study also showed that participants with limited prior knowledge of physical exercise and limited personal resources to develop strategies face greater challenges in implementing strategies for maintaining physical activity.

This study demonstrated that health-related challenges, including comorbidities, short-term illnesses, and stroke-related symptoms, were substantial barriers to maintaining physical activity after the mHealth intervention. The impact of comorbidities is an especially important consideration for health promotion interventions targeting individuals poststroke or TIA, as comorbidities are more prevalent in this group than in the general population and have a substantial negative influence on both functional outcomes and mortality [43,44]. In addition, comorbidity is strongly linked to age, with most adults older than 65 years having one or more diagnoses, and because people

XSL•FO
RenderX

poststroke or TIA are often older, age-related comorbidities are frequent [45]. Furthermore, consistent with prior research [36,46,47], the participants struggled to allocate time for physical activity along with other responsibilities and engagements. These findings underscore that maintaining physical activity could be challenging for people poststroke or TIA given the outlined health-related and contextual barriers. Hence, these circumstances should be carefully considered in the design and development of mHealth interventions. Building on this study's findings that personal strategies can help overcome health-related and contextual barriers, mHealth interventions should include components that support the implementation of such strategies in complex life situations. These components should balance the reinforcement and adaptation of existing strategies among experienced and resourceful participants, with targeted support for those who face greater challenges in the implementation.

Implications

Our results highlight the important consideration of the interaction between individuals and their environment while supporting the transition from therapeutic support to self-management in mHealth physical activity interventions for people poststroke or TIA. One potential approach is to include a structured maintenance phase in the intervention design, during which therapeutic support is gradually reduced. In this phase, participants can progressively take more responsibility for their own development and independently apply self-regulation techniques to sustain physical activity. Because behavior maintenance is a dynamic process, discontinuities are to be expected; therefore, training participants to manage setbacks is recommended [39], including strategies that foster agency and resilience to overcome health-related and contextual barriers. Our findings further indicate that support from significant others, through encouragement and companionship, is essential for maintaining physical activity. Accordingly, future mHealth interventions should be aware of the complex situation for this population and help participants identify personal strategies in daily life and engage with a supportive social environment beyond the formal intervention setting. Finally, the study findings should be interpreted within the Nordic context, as, for instance, different climates, seasonal contexts, or digital health care systems may yield distinct barriers and facilitators.

Strengths and Limitations

A major strength for the study's trustworthiness was the design and methodology, which contributed to the rich dataset developed through repeated semistructured interviews at 3 months and 6 months postintervention and participant-generated photographs during the second interview. This approach facilitates going beyond surface-level descriptions of participants' experiences by building rapport and stimulating reflection between and during sessions [23,34,48]. Further, using participant-generated photos may help reduce power imbalances between researcher and participant while also enabling the introduction of new dimensions to data generation that might not be anticipated in the interview guide, thereby contributing to a deeper understanding of the issue under study [23,34]. Furthermore, participant-generated photos and interviews at 3 months and 6 months contributed to reducing recall and response bias and to gaining both short- and long-term experience with maintaining physical activity. Additional follow-up could have provided more insight into maintaining physical activity. The design was grounded in a constructivist epistemology, with reflexivity as a central component in the generation and analysis of data [26,27]. The first author had brief contact with the participants during the feasibility trial's data collection phase. This dual role may induce the risk of social desirability bias. In addition to the design with repeated interviews and participant-generated photos, several strategies were used to elicit authentic interview responses, such as providing assurances and probing questions [49]. These strategies for an authentic and rich dataset were developed through reflexive discussions of pilot interviews, followed by continuous reflexivity through data generation.

Limitations of the study include a lack of diversity in some of the aspects of the sample recruited from the mHealth version of the i-REBOUND intervention [21]. Diversity was obtained in terms of geography, age, educational background, and experiences with daily use of the internet and mobile apps. However, the relatively high physical activity and functional level, gender imbalance, and lack of a more complete socioeconomic background represent limitations that influence the transferability of our findings. The challenges with recruiting a diverse sample were previously addressed [21,50]. Although the participants in this study were a relatively homogeneous group, they still faced a variety of barriers related to maintaining physical activity after the mHealth intervention and needed support to adjust their physical activity to meet recommendations given their current life situations. However, our findings show important facilitators, such as knowledge acquisition and support from the social and environmental context, that can be useful for exploring conditions for more diverse populations. To facilitate the judgment of transferability to other contexts and populations, we reported detailed descriptions of the study's methodology [27,51]. To explore how gendered and socioeconomic aspects as well as experiences of those with a more sedentary lifestyle and severe stroke influence physical activity maintenance, recruitment strategies in future studies should aim for a more diverse sample, not the least because socioeconomic disparities are shown to be related to stroke incidence and outcome [52].

Conclusions

This study points out the need for future mHealth interventions to better prepare people poststroke and TIA to develop and apply personalized strategies for managing barriers to maintaining physical activity. The complex life situation following stroke or TIA, involving health-related challenges and contextual barriers, can limit individuals' ability to engage in regular physical activity. Nevertheless, participants may develop strategies, supported by knowledge and tools from mHealth interventions and the individuals' surrounding context, to navigate these challenges. These insights highlight opportunities for future mHealth interventions to strengthen the interaction between individuals and their environment while also empowering personalized strategies for supporting behavior change and long-term physical activity maintenance.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all participants for their valuable contribution to this study. To assist with language refinement and editorial improvements during manuscript writing, the generative artificial intelligence tool Microsoft Copilot was used. All intellectual content, including the study's design, interpretation, and conclusions, is entirely the authors' own.

Funding

This work was supported by grants from the Swedish Research Council (2022-01403); Swedish Research Council for Health, Working Life and Welfare (2021-01018); Center for Innovative Medicine, Region Stockholm (FoUI-960631); Sweden's Innovation Agency (2021-01726); Swedish Stroke Association; Strategic Research Area Health Care Sciences at Karolinska Institutet; and Research School of Health Science at Karolinska Institutet.

Data Availability

The dataset used in this article is not publicly available due to the risk of compromising research participants' privacy in accordance with Swedish legislation. Requests to access data should be directed to the Research Data Office at Karolinska Institutet (rdo@ki.se).

Authors' Contributions

Conceptualization: HL, DMC, CKH, LK, LH
Data Curation: HL, LH
Formal Analysis: HL, DMC, CKH, LK, LH
Funding Acquisition: DMC
Investigation (interviewing): HL
Methodology: HL, LH
Resources: DMC
Project Administration: HL
Writing – Original Draft: HL, LH
Writing – Review & Editing: HL, DMC, CKH, LK, LH

Conflicts of Interest

None declared.

References

1. Feigin VL, Stark BA, Johnson CO, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820. [doi: 10.1016/S1474-4422(21)00252-0]
2. Hankey GJ. Secondary stroke prevention. Lancet Neurol 2014 Feb;13(2):178-194. [doi: 10.1016/S1474-4422(13)70255-2]
3. Boss HM, Deijle IA, Van Schaik SM, et al. Cardiorespiratory fitness after transient ischemic attack and minor ischemic stroke: baseline data of the MoveIT study. J Stroke Cerebrovasc Dis 2017 May;26(5):1114-1120. [doi: 10.1016/j.jstrokecerebrovasdis.2016.12.029] [Medline: 28089564]
4. Pasztor B, Drummond A, Hancox JE, Taylor IM. A systematic scoping review of poststroke engagement in physical activity. Arch Rehabil Res Clin Transl 2025 Jun;7(2):100457. [doi: 10.1016/j.arrct.2025.100457] [Medline: 40678284]
5. McEwan D, Rhodes RE, Beauchamp MR. What happens when the party is over?: Sustaining physical activity behaviors after intervention cessation. Behav Med 2022;48(1):1-9. [doi: 10.1080/08964289.2020.1750335] [Medline: 32275199]
6. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
7. Dasso NA. How is exercise different from physical activity? A concept analysis. Nurs Forum 2019 Jan;54(1):45-52. [doi: 10.1111/nuf.12296] [Medline: 30332516]
8. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131. [Medline: 3920711]
9. Laver KE, Adey-Wakeling Z, Crotty M, Lannin NA, George S, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020 Jan 31;1(1):CD010255. [doi: 10.1002/14651858.CD010255.pub3] [Medline: 32002991]
10. Allan LP, Beilei L, Cameron J, et al. A scoping review of mHealth interventions for secondary prevention of stroke: implications for policy and practice. Stroke 2023 Nov;54(11):2935-2945. [doi: 10.1161/STROKEAHA.123.043794] [Medline: 37800373]

XSL•FO
RenderX

11. Rintala A, Kossi O, Bonnechère B, Evers L, Printemps E, Feys P. Mobile health applications for improving physical function, physical activity, and quality of life in stroke survivors: a systematic review. Disabil Rehabil 2023 Dec;45(24):4001-4015. [doi: 10.1080/09638288.2022.2140844] [Medline: 36325613]
12. Thurston C, Johansson S, von Koch L, English C, Conradsson DM. Cross-sector collaboration to develop a mobile health solution for promotion of physical activity after stroke or transient ischaemic attack in Sweden. BMC Digit Health 2025;3(1):1-13. [doi: 10.1186/s44247-025-00168-2]
13. Hall P, Lawrence M, Blake C, Lennon O. Interventions for behaviour change and self-management of risk in stroke secondary prevention: an overview of reviews. Cerebrovasc Dis 2024;53(1):1-13. [doi: 10.1159/000531138] [Medline: 37231867]
14. Dunton GF, Leventhal AM, Rebar AL, Gardner B, Intille SS, Rothman AJ. Towards consensus in conceptualizing and operationalizing physical activity maintenance. Psychol Sport Exerc 2022 Jul;61:102214. [doi: 10.1016/j.psychsport.2022.102214]
15. Nicholson SL, Donaghy M, Johnston M, et al. A qualitative theory guided analysis of stroke survivors' perceived barriers and facilitators to physical activity. Disabil Rehabil 2014;36(22):1857-1868. [doi: 10.3109/09638288.2013.874506] [Medline: 24377329]
16. Sammut M, Haracz K, Shakespeare D, et al. Physical activity after transient ischemic attack or mild stroke is business as usual. J Neurol Phys Ther 2022 Jul 1;46(3):189-197. [doi: 10.1097/NPT.0000000000000395] [Medline: 35727994]
17. Madigan CD, Grimmett C, Sweet SN, et al. Understanding adult's experiences and perceptions of how to maintain physical activity: a aystematic review and qualitative synthesis. Int J Behav Med 2025 Oct;32(5):687-701. [doi: 10.1007/s12529-024-10335-w] [Medline: 39557793]
18. Huffman MK, Amireault S. What keeps them going, and what gets them back? Older adults' beliefs about physical activity maintenance. Gerontologist 2021 Apr 3;61(3):392-402. [doi: 10.1093/geront/gnaa087] [Medline: 32622348]
19. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory: Prentice-Hall; 1986.
20. English C, Ramage ER, Attia J, et al. Secondary prevention of stroke. A telehealth-delivered physical activity and diet pilot randomized trial (ENAbLE-pilot). Int J Stroke 2024 Feb;19(2):199-208. [doi: 10.1177/17474930231201360] [Medline: 37658738]
21. Thurston C, Humphries S, Bezuidenhout L, et al. Mobile health delivered physical activity after mild stroke or transient ischemic attack: is it feasible and acceptable? Int J Stroke 2025 Aug;20(7):801-811. [doi: 10.1177/17474930251315628] [Medline: 39800974]
22. Bezuidenhout L, Humphries S, English C, Sundberg CJ, Nilsson M, Conradsson DM. Physiotherapists' delivery of a mobile health physical activity intervention for people post-stroke and transient ischemic attack: insights from a feasibility study. Digit Health 2025;11:20552076251374247. [doi: 10.1177/20552076251374247] [Medline: 40918082]
23. Harper D. Talking about pictures: a case for photo elicitation. Visual Studies 2002 Apr;17(1):13-26. [doi: 10.1080/14725860220137345]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
26. Braun V, Clarke V. Thematic Analysis: A Practical Guide: SAGE; 2022.
27. Braun V, Clarke V. Supporting best practice in reflexive thematic analysis reporting in Palliative Medicine: a review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). Palliat Med 2024 Jun;38(6):608-616. [doi: 10.1177/02692163241234800]
28. Thurston C, Bezuidenhout L, Humphries S, et al. Mobile health to promote physical activity in people post stroke or transient ischemic attack - study protocol for a feasibility randomised controlled trial. BMC Neurol 2023 Mar 28;23(1):124. [doi: 10.1186/s12883-023-03163-0] [Medline: 36978045]
29. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462. [doi: 10.1136/bjsports-2020-102955]
30. Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: what's in a name? A review of literature. Eur J Gen Pract 1996;2(2):65-70. [doi: 10.3109/13814789609162146]
31. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke 2007 Mar;38(3):1091-1096. [doi: 10.1161/01.STR.0000258355.23810.c6] [Medline: 17272767]
32. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The Fatigue Severity Scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989 Oct;46(10):1121-1123. [doi: 10.1001/archneur.1989.00520460115022] [Medline: 2803071]
33. Green EM, Spivak C, Dollahite JS. Early adolescent food routines: a photo-elicitation study. Appetite 2021 Mar 1;158(158):105012. [doi: 10.1016/j.appet.2020.105012] [Medline: 33132192]
34. Bates EA, McCann JJ, Kaye LK, Taylor JC. "Beyond words": a researcher's guide to using photo elicitation in psychology. Qual Res Psychol 2017 Oct 2;14(4):459-481. [doi: 10.1080/14780887.2017.1359352]
35. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]

XSL•FO
RenderX

36. Wahlich C, Beighton C, Victor C, et al. “You started something … then I continued by myself”: a qualitative study of physical activity maintenance. Prim Health Care Res Dev 2017 Nov;18(6):574-590. [doi: 10.1017/S1463423617000433] [Medline: 28774352]
37. Mantey DS, Hunt ET, Hoelscher DM, Kelder SH. Social cognitive theory and health behaviour. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice, 6th edition: Jossey-Bass; 2024.
38. Deijle IA, Van Schaik SM, Van Wegen EEH, Weinstein HC, Kwakkel G, Van den Berg-Vos RM. Lifestyle interventions to prevent cardiovascular events after stroke and transient ischemic attack: systematic review and meta-analysis. Stroke 2017 Jan;48(1):174-179. [doi: 10.1161/STROKEAHA.116.013794] [Medline: 27924055]
39. Bandura A. Self-Efficacy: The Exercise of Control: W. H. Freeman; 1997.
40. Lagerlund H, Bezuidenhout L, Humphries S, et al. Adherence to and engagement with an mHealth physical activity intervention after mild stroke or transient ischemic attack: secondary analysis of a feasibility randomized controlled trial. JMIR Mhealth Uhealth 2026;14:e75662-e75662. [doi: 10.2196/75662]
41. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020;22(12):e21687. [doi: 10.2196/21687]
42. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
43. Gallacher KI, Batty GD, McLean G, et al. Stroke, multimorbidity and polypharmacy in a nationally representative sample of 1,424,378 patients in Scotland: implications for treatment burden. BMC Med 2014 Oct 3;12(1):151. [doi: 10.1186/s12916-014-0151-0] [Medline: 25280748]
44. Sennfält S, Pihlsgård M, Petersson J, Norrving B, Ullberg T. Long-term outcome after ischemic stroke in relation to comorbidity - an observational study from the Swedish Stroke Register (Riksstroke). Eur Stroke J 2020 Mar;5(1):36-46. [doi: 10.1177/2396987319883154] [Medline: 32232168]
45. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
46. Jansons PS, Robins L, Haines TP, O’Brien L. Barriers and enablers to ongoing exercise for people with chronic health conditions: participants’ perspectives following a randomized controlled trial of two interventions. Arch Gerontol Geriatr 2018;76:92-99. [doi: 10.1016/j.archger.2018.02.010] [Medline: 29477950]
47. Maula A, LaFond N, Orton E, et al. Use it or lose it: a qualitative study of the maintenance of physical activity in older adults. BMC Geriatr 2019 Dec 12;19(1):349. [doi: 10.1186/s12877-019-1366-x] [Medline: 31830900]
48. Polkinghorne DE. Language and meaning: data collection in qualitative research. J Couns Psychol 2005;52(2):137-145. [doi: 10.1037/0022-0167.52.2.137]
49. Bergen N, Labonté R. “Everything is perfect, and we have no problems”: detecting and limiting social desirability bias in qualitative research. Qual Health Res 2020 Apr;30(5):783-792. [doi: 10.1177/1049732319889354] [Medline: 31830860]
50. Johnsson C, Asaba E, Guidetti S, Åkesson E, Hagströmer M, Patomella AH. Make my day - stroke prevention grounded in engaging everyday activities in primary healthcare - a single-blinded randomised controlled trial. J Prim Care Community Health 2025;16:21501319251385889. [doi: 10.1177/21501319251385889] [Medline: 41178128]
51. Williams EN, Morrow SL. Achieving trustworthiness in qualitative research: a pan-paradigmatic perspective. Psychother Res 2009 Jul;19(4-5):576-582. [doi: 10.1080/10503300802702113] [Medline: 19579089]
52. Pantoja-Ruiz C, Akinyemi R, Lucumi-Cuesta DI, et al. Socioeconomic status and stroke: a review of the latest evidence on inequalities and their drivers. Stroke 2025 Mar;56(3):794-805. [doi: 10.1161/STROKEAHA.124.049474] [Medline: 39697175]

Abbreviations

mHealth: mobile health
SCT: social cognitive theory
STAAR: Stroke Treatment through Active and Accessible Rehabilitation
TIA: transient ischemic attack

XSL•FO
RenderX

Edited by A Coristine; submitted 18.Nov.2025; peer-reviewed by F Kennedy; revised version received 29.Apr.2026; accepted 29.Apr.2026; published 27.May.2026.

Please cite as:
Lagerlund H, Moulaee Conradsson D, Häger CK, Kwak L, Holmlund L
Barriers and Facilitators to Maintaining Physical Activity After an mHealth Intervention for People Poststroke or Transient Ischemic Attack: Reflexive Thematic Analysis
JMIR Cardio 2026;10:e87493
URL: https://cardio.jmir.org/2026/1/e87493
doi:10.2196/87493

© Hanna Lagerlund, David Moulaee Conradsson, Charlotte K Häger, Lydia Kwak, Lisa Holmlund. Originally published in JMIR Cardio (https://cardio.jmir.org), 27.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Experiences of Patients With Atrial Fibrillation Using Technology to Personalize Self-Care Decision-Making: Interpretive Description Study

Kathy L Rush[1], PhD, RN; Cherisse L Seaton[1], PhD; Alexandra Lukey[1], PhD, RN; Victor Wang[1]; Peter S Loewen[2], PharmD; Wynne Chiu[3], MSN, CCN(c), RN; Megan Patapoff[4], MN-NP(f); Lana Moroz[5], NP; Jason G Andrade[5], MD, FRCPC

[1]School of Nursing, Faculty of Health and Social Development, The University of British Columbia, 1147 Research Rd,, Kelowna, BC, Canada

[2]Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, BC, Canada

[3]Heart Failure and Transplantation, St Paul's Hospital, Vancouver, BC, Canada

[4]Ambulatory Cardiology, Vancouver General Hospital, Vancouver, BC, Canada

[5]Cardiac Atrial Fibrillation Clinic at Diamond Health Centre, Vancouver, BC, Canada

Corresponding Author:
Kathy L Rush, PhD, RN
School of Nursing, Faculty of Health and Social Development, The University of British Columbia, 1147 Research Rd,, Kelowna, BC, Canada

Abstract

Background: Atrial fibrillation (AF) is the most common sustained heart rhythm disorder and is a challenging chronic disease to manage. Patients' daily self-care decisions are associated with improved AF outcomes, quality of life, and decreased hospital use and cost. However, many patients find these real-world or naturalistic decisions difficult, often because of their inherent complexity and ambiguity, coupled with the uncertainty of AF. Intervention research using technology to support AF self-care has largely emphasized making decisions with clinicians. Patients with AF are increasingly using consumer technology; yet, little is known about the use of technology by patients with AF in independent self-care decision-making. Addressing this gap will facilitate developing interventions that better leverage technology to enhance patients' naturalistic decision-making.

Objective: This study aimed to explore the experiences of older adult patients in using technology to support self-care decision-making.

Methods: Following an interpretive descriptive qualitative approach, older adult patients with AF were recruited from 3 specialty heart function clinics in a Western Canadian province to participate in 1 of 6 facilitated virtual focus groups for 1.5 hours. Patients were asked about their self-care decision-making since AF diagnosis, their AF-specific technology use and its use in making self-care decisions, their technology motivations, benefits, constraints, and other possibilities for use. Inductive thematic analysis was used to code the transcribed data, moving from open coding to clustering of common codes into categories, looking for patterns of meaning between and across categories to iteratively arrive at main themes and subthemes.

Results: Thirty patients (n=15, 50% women) with AF (mean age, 73, SD 5.7 years; range 63 - 85 years) participated in the focus groups. Participants' experiences of using technology to make daily self-care decisions were highly variable but centered on its personalized use to meet their individualized needs, preferences, and life context. The personalizing process of technology use in decision-making was characterized by three themes: (1) beginning technology use in their own times and ways, during their AF trajectory—pre-, at the time, at some point AF post diagnosis, and could be either self-initiated and/or provider recommended or influenced; (2) developing patterns of AF self-care decision-making using technology, including establishing their personal baseline, keeping out of the danger zone, watchful waiting, and seeking decision-making support; and (3) finding the place for technology in normalizing daily life, either settling on or limiting its use to normalize life.

Conclusions: Findings expand understandings of naturalistic decision-making by elucidating the personalized process of technology use in AF self-care.

(JMIR Cardio 2026;10:e93036) doi:10.2196/93036

KEYWORDS

cardiovascular; technology; self-care; experience; qualitative; naturalistic decision-making; atrial fibrillation; mobile phone

XSL•FO
RenderX

Introduction

Atrial fibrillation (AF) is the most commonly sustained heart rhythm disorder in older adults and is a challenging chronic disease to manage. Self-care, defined as a decision-making process for managing illness and maintaining health through health-promoting practices, is a cornerstone of AF management [1]. Despite self-care's demonstrable improvement in AF outcomes (eg, reduced complications such as stroke), quality of life, reduced health care usage and cost, and decreased treatment and relapse (eg, ablation) [2], patients with AF have found self-care decisions challenging to make [3,4]. Although providers play an important role in patients' treatment decision-making and are a source of reassurance [5], patients spend only about 0.001% (10 hours) of their time per year with health care providers [6]. In their daily lives, patients are faced with many other decisions not directly related to treatment, such as activity, risk behaviors, and health care access [7]. Occurring in complex, real-world settings under uncertainty and ambiguity, such decisions are referred to as naturalistic decision-making.

Naturalistic decision-making has appeared in the literature since the 1980s [8]; however, its entrance into health care and cardiovascular research has been limited. Notably, research on naturalistic decision-making has demonstrated that people do not generate and compare options in systematic ways when making decisions in real-world contexts [8]; instead, they rely on past experience to efficiently categorize situations, prototypical of the Recognition-Primed Decision Model, which emphasizes situation awareness and simulating options for a course of action [9]. Similarly, with respect to cardiovascular disease self-care decisions related to medication, pain management, diet, and exercise, older adult patients did not always follow a rules-based approach in compliance with instructions from providers; instead, patients often relied on their knowledge and experience when making decisions, such as not walking when it is cold or windy, since it might affect them later on [10]. Although not specific to AF, 2 studies have explored the naturalistic decision-making processes among patients with heart failure. First, in a reanalysis of interview data, Riegel et al [11] confirmed that situation awareness (ie, recognition of a symptom and assessment of severity or importance) and mental simulation of action (ie, exploring and developing plans of action considering outcomes) were key processes underlying self-care decisions. Second, Daley et al [12] further detailed how naturalistic self-care decision-making occurred in both sequential and independent phases of monitoring, interpreting, and acting, but did not specifically study the role of technology in these processes. The use of technology in AF care has evolved rapidly and become ubiquitous, evident in the growing number of clinical trials using digital health technology for patients with AF [13]. Intervention research using digital health technologies to support self-care has primarily been clinician-driven and often embedded within integrated management systems for remote monitoring of patients' biometric data [14,15]. For example, in a recent systematic review of AF self-care interventions, mobile and web-based interventions combined with strategies such as personalized education and continuous support showed promise in enhancing patient-reported clinical and health care utilization outcomes [15]. Despite a handful of intervention studies illustrating through qualitative follow-up methods how technology helped patients in their self-management of AF symptoms, lifestyle, and anxiety [16-19], these digital interventions were dependent on health providers, whose support, while important, may reinforce patients' dependence on their providers for their self-care.

Independent of the health care service context, consumer technologies for heart monitoring and tracking are gaining popularity among patients with AF. In a survey administered through StopAF.org, 71% of respondents reported using consumer technology to manage their AF, most commonly KardiaMobile devices and Apple smartwatches [20]. A recent qualitative study of patients with AF using self-tracking consumer technologies reported that patients used devices to provide confirmation and reassurance of their bodily experiences and facilitate greater readiness for shared decision-making with providers [21]; however, naturalistic decision-making was not the focus of the study and little is known about how patients with AF use technology to make daily self-care decisions independently of their health care providers. Self-care is highly disease-specific [22], and naturalistic decision-making among patients with AF has been understudied. Therefore, the purpose of this study was to explore the experiences of older adult patients with AF in using technology to support self-care decision-making.

Methods

Study Design

This study used an interpretive description design [23], which is aligned with a constructivist and naturalistic orientation to inquiry. It is appropriate for exploratory, yet clinically applicable, phenomena such as patients' self-care decision-making experiences using technology and the contexts in which it occurs. Reporting follows the COREQ (Consolidated Criteria for Reporting Qualitative Research) [24].

Ethical Considerations

The study was approved by the Behavioral Research Ethics Board at The University of British Columbia (#H24-02365). All participants provided informed consent. Study data were deidentified and participant numbers were used to replace participant names so that no individual participants could be identified. Participants were offered a CAD $50 (CAD $50=approximately US $36, at the time of the study) eGift certificate for participation.

Sampling, Inclusion, and Recruitment

Participants were recruited from 3 outpatient cardiac clinics in a Western Canadian province in the 2 months prior to focus groups using in-person recruitment, email, or telephone calls. Patient eligibility included having AF, aged 60 years or older following the World Health Organization definition of older adults [25], able to provide informed consent, able to understand English, and able to join one of the scheduled 90-minute focus groups by Zoom (Zoom Communications, Inc) or telephone. Technology use was defined broadly to ensure inclusivity of

XSL•FO
RenderX

all possible sources patients might use in their self-care decision-making, and patients were given as an example "use of the internet" for this study's purpose. Patients with self-reported limited technology use were not excluded but encouraged to join. Patients were screened from daily patient clinic lists and eligible patients either met with a research assistant in clinic prior to an appointment or were telephoned. Participants were given a high-level overview of the research and more information about the study from the consent form.

Data Collection

Focus groups, held in June and July 2025 using the Zoom videoconference software (with telephone option), were used to elicit participants' experiences and insights about their technology use in self-care decision-making. The virtual approach facilitated patient participation by avoiding travel logistics and costs. Before the focus groups, all participants provided consent and completed a short online survey (paper versions were offered), including demographic, cardiovascular disease, and technology-related questions. To give facilitators a sense of the technology use and confidence of participants attending the sessions, participants were asked to self-select whether they considered themselves a high or low technology user, and to rate their confidence in using technologies on a single item 3-point scale ranging from *not very confident* (1) to *very confident* (3) [26]. Participants were asked to select which of 3 devices they used (computer, smartphone, and tablet) in general and also select from a researcher-generated list of 8 digital technologies specific to cardiovascular disease management or indicate "I don't use any."

Focus groups, which were approximately 1.5 hours in length, were guided by a semistructured question guide asking about participants' experiences with technologies and their use in day-to-day health and self-care decisions, with prompts used to encourage elaboration or clarification as needed (Textbox 1). Questions were broadly guided by the conceptualization by Riegel et al [11] of self-care as a naturalistic decision-making process in self-maintenance, monitoring, and managing, while integrating the role of technology. Two patient partner (1 male and 1 female) team members reviewed the proposed questions and provided feedback to simplify language. Focus groups were zoom audio-recorded and facilitated by experienced team members (KLR [PhD and female professor] and a male master's prepared non–team member working as a research coordinator for KLR on other aging-related projects who supported data collection), both of whom had no prior relationship with participants. Other team members provided technical support and/or took notes during the session. The focus groups were fairly evenly distributed in terms of demographic and technology use composition. "Code" saturation, defined as the point when no additional codes are identified in the data, was reached at focus group 3 and "meaning" saturation, defined as the point when no further insights or nuances within the codes are identified, was reached at focus group 6 [27].

XSL•FO
RenderX

Textbox 1. Focus group guiding questions and prompts.

Self-care decision-making and the role of technology

1. "Reflecting on when you first received a diagnosis to present day, how have you made everyday decisions (meaning decisions that you have to make everyday about your condition) to manage your cardiovascular health and care?"
2. "How you are currently using digital technologies, meaning any type of modern device or technology, to manage your cardiovascular health and care. This includes things like smartwatches, medication reminders, online lab results, apps, or even just using the Internet to find health information. Can you tell us about any ways you are using digital technologies?"

   Possible follow-up probes (if not discussed)

   - "Specifically, can you first tell us about any ways you are using digital technologies to help you maintain your health (to support things like safe exercise or reducing risks)?"
   - "Secondly, how about using digital technologies for cardiovascular monitoring such as checking for changes in signs and symptoms (things like checking heart rate and rhythm, pulse, and blood pressure monitoring)?"
   - "Thirdly, can you tell us about how you are currently using digital technologies to manage any changes (e.g., determining treatment changes and decision making about when to seek care)?"

3. "How do you use information you are collecting from your technology to make decisions about your self-care? Probe: Can you talk about an incident using technology where you had to make a decision about your self-management? What do you do with all the information that you collect (e.g., are your providers interested in the information you are collecting)?"
4. "What motivates you to collect that data, and to use the technologies in the way that you are to look after your cardiovascular care (i.e., to maintain, monitor, and manage your cardiovascular care)? Probe: was it cardiovascular disease prompted, or a technology you used prior? Probe: who suggested the technologies—a provider or spouse or was it your independent decision?"
5. "Please talk about any benefits you have found with the technologies? What about barriers to accessing/using them—Are there any constraints that have maybe kept you from using technologies to the extent that you might or could?"
6. "Whether you are using technology or not can you imagine possibilities for using digital technologies for self-care in new or different ways than you currently are? In what ways have digital technologies met your needs in managing your cardiovascular care? Where could technology still be improved in terms of meeting your needs?"

Summary and conclusion

1. "Is there anything else related to digital technology use in cardiovascular disease self-management that you would like the team to know?"
2. "Do you have any final comments or questions in relation to this session today or this research project before we let you go?"

Data Analysis

Participant responses to the initial survey were analyzed descriptively. Focus group recordings were transcribed automatically using Zoom software and accuracy checked against the recording by an undergraduate student research assistant (VW). Research team members listened to the audio recordings of the focus groups and read through the focus group transcripts to familiarize and immerse themselves in the data and engaged in ongoing reflexivity and reflection to promote rigor, sincerity, and credibility [28]. The approach of Richards and Morse [29] guided inductive thematic analysis moving from open coding to cluster common codes into categories, looking for patterns of meaning between and across categories to iteratively arrive at main themes and subthemes (Multimedia Appendix 1). Two team members (AL and KLR) analyzed the data, initially coding for units of meaning (eg, words, phrases, or paragraphs) across groups to construct an initial coding schema and following discussion analyzing for meaning across and within clusters of codes and refining and modifying to iteratively create themes and subthemes. Coding was completed using NVivo (version 12; Lumivero) [30].

Results

Participant Characteristics

Six focus groups were conducted with 5 - 6 participants in each (total N=30), with 3 participants joining via the telephone option. The mean age of the participants was 73.4 (SD 5.7) years, with 15 (50%) identifying as women. All participants had at least 1 device, most commonly a smartphone. Two-thirds used a wearable and/or heart tracker of some kind, and half used online laboratory services (see Table 1 for full participant characteristic details).

XSL•FO
RenderX

Table . Characteristics of older adult focus group participants with atrial fibrillation.

Characteristic	All focus group participants (N=30)
Age (range 63-85) years, n (%)	
60 - 69	7 (23.3)
70 - 79	18 (60)
80+	5 (16.7)
Gender, n (%)	
Woman	15 (50)
Man	15 (50)
Ethnicity, n (%)	
European/European-Canadian	20 (66.7)
Middle Eastern	2 (6.7)
European+Indigenous	2 (6.7)
Hispanic/Latino	1 (3.3)
Indigenous (First Nations, Métis)	1 (3.3)
Prefer not to answer	1 (3.3)
Other (eg, “Canadian”)	3 (10)
Highest education level, n (%)	
University degree	19 (63.3)
Some college or university	8 (26.7)
Trades certificate/college diploma	2 (6.7)
Prefer not to answer	1 (3.3)
Annual household income, n (%)	
Over $100,000 (US $72,400)	9 (30)
$75,000 to $99,999 (US $54,300 to US $72,399)	4 (13.3)
$50,000 to $74,999 (US $36,200 to US $54,299)	2 (6.7)
$25,000 to $49,999 (US $18,100 to US $36,199)	5 (16.7)
Less than $24,999 (US $18,099)	2 (6.7)
Prefer not to answer	8 (26.7)
Marital status, n (%)	
Married/remarried	21 (70)
Divorced	4 (13.3)
Widowed	3 (10)
Single	2 (6.7)
High or low technology use (self-ranked), n (%)	
High technology user	17 (56.7)
Low technology user	12 (40)
Missing/not reported	1 (3.3)
Self-confidence in using digital technologies, n (%)	
Very confident	9 (30)
Somewhat confident	19 (63.3)
Not very confident	2 (6.7)
Device use (multiple answers), n (%)	
Smartphone	28 (93.3)

XSL•FO
RenderX

Characteristic	All focus group participants (N=30)
PC or Apple computer	26 (86.7)
Tablet	20 (66.7)
I don’t use any devices	0 (0)
Presence of cardiovascular disease (multiple answers), n (%)	
Atrial fibrillation/atrial flutter	30 (100)
Hypertension	9 (30)
Heart failure	4 (13.3)
Digital health technologies used in cardiovascular disease management (multiple answers), n (%)	
Online laboratory results/services	15 (50)
Online appointment booking	14 (46.7)
Wearable devices	14 (46.7)
Heart trackers/monitors	8 (26.7)
Smartphones/tablet applications	6 (20)
Online prescription refills	6 (20)
Video calls	5 (16.7)
Patient portal	1 (3.3)
Other (eg, “Holter”; “blood pressure monitor”)	4 (13.3)

During the focus group discussions, participants described their use of different technologies in AF self-care decision-making, with some avid technology users and others more limited. Common uses were technology to track biometrics, to search the internet for general health and AF-specific information, and to access health reports such as laboratory results online.

Themes

Patient participants’ variable experiences of using technology to make self-care decisions were characterized by a personalizing process in which they tailored its use according to their unique needs, preferences, and life context. Three themes described this personalizing process across patients: (1) beginning technology use in their own times and ways, (2) developing patterns of AF self-care decision-making using technology, and (3) finding the place for technology in normalizing daily life. Each theme was defined by subthemes and participants’ perspectives, emotions, and concerns related to the process.

Beginning Technology Use in Their Own Times and Ways

The process of personalizing technology use began for participants at different times and in different ways during their AF trajectory—pre-, at the time, and at some point of AF post diagnosis. It began with participants taking the initiative and/or because of a provider’s recommendation or influence.

Participant-Initiated Beginnings

An AF event often catalyzed participants’ decisions to initiate technology use in their self-care. Prediagnostically, validating troublesome symptoms (eg, feeling unwell and nearly passing out) using technology was often the deciding prompt such as 1 participant who purchased their own smartwatch immediately after using a friend’s watch that showed AF on an electrocardiogram (ECG) and confirming with their doctor (Participant 8).

An AF diagnosis was the impetus for several participants to initiate technology use, either acquiring new technology for the first time, adding AF-specific technology (eg, KardiaMobile) to what they already owned, or using their current technology for their AF self-care. AF was the sole reason for several participants to purchase technology, such as Participant 8 who admitted, “But the reason I wear it [smartwatch]- I never would have bought it if it hadn’t been for the atrial fibrillation” while Participant 26 echoed having “no need for a smartwatch prior to [AF diagnosis]” but described it as “a [game] changer for me.” Another participant described blending their current with newly acquired technology “brought on by my heart condition” (Participant 28).

Provider Recommendations and Influences

Multiple participants described the influence of their physician’s recommendations or perceptions of the technology (eg, tracking device) in their decision to begin and/or continue using technology. Participants described providers giving them practical help to maximize the use of their existing technology (Apple Watch to run ECG) or asking them to “run a couple of strips and email them to me if you ever have issues” (Participant 23), or to have copies of the data for their records. However, not all participants responded as positively to provider recommendations to use monitoring devices, with some expressing greater reluctance as one participant described, “I’ve

XSL•FO
RenderX

been recommended to use all those things, and I'm pretty slack at it, but it's a matter of getting used to doing it all the time, and not begrudging the fact that you have to" (Participant 16).

In contrast, a number of participants described providers who were less receptive and interested in their technology data, which influenced their decision-making. Lack of provider interest led 1 participant to decide against paying the additional costs to upgrade and connect their KardiaMobile to the internet to transmit data (Participant 9). One participant, who used only a step counter app for measuring walking steps and no other device for AF monitoring noted that "my doctor hasn't really suggested any devices so far" (Participant 20).

Even when doctors recommended technology use, less technologically savvy patients were frustrated by their lack of digital literacy either not using it because they don't "know how to work the thing" (eg, KardiaMobile) (Participant 4) or delaying device adoption until their AF became uncontrolled, as one participant elaborated,

> *My GP [general practitioner] and my cardiologist wanted me to have an Apple Watch. And I thought, oh great, just another piece of technology that I will not understand. So eventually my AFib got out of control, and so I marched myself over to the Apple store.* [Participant 25]

Developing Patterns of AF Self-Care Decision-Making Using Technology

Overview

Regardless of when participants initiated their use of technology, they described using 1 or more devices singly or in combination for making daily self-care decisions. Participants personalized their use of technology in their AF self-care decision-making based on several contextual factors: type of AF (symptomatic or asymptomatic), daily activity, technology preferences, comfort levels, data relevance to their situation, and whether responding to specific AF episodes or proactively to mitigate AF exacerbations and their consequences. Overall, participants were sophisticated active data managers and processors evident in their use of 1 or more of 4 technology-driven decision-making patterns: (1) establishing their personal baseline for decision-making, (2) keeping out of the danger zone, (3) watchful waiting, and (4) seeking decision-making support. The ability to "self-diagnose" and confirm episodes of AF through technology was lauded as a way of gaining some control, when in the past participants had to wait for information from their providers, "It's really great to have any kind of diagnostic tool, just to have a feeling of empowerment and learning more about your own condition, instead of being in the dark as way back in the past" (Participant 12).

Establishing Their Personal Baseline for Decision-Making

Participants found that technology facilitated learning more about themselves in the context of their AF to personalize their self-care decision-making. Despite variability in the consistency, frequency, type, amount, and sources of data they gathered, core to participants' self-care decision-making was establishing their baseline. Participants consistently referred to using their data to "benchmark," "trend," "pattern," and "compare" that was foundational to determining changes in their condition that informed daily decision-making.

Several participants did very extensive tracking, trending, and analysis of their AF data over time, using their preferred method of either paper or devices/apps. Some participants had been doing this for several years—one since their cardiologist's recommendation 20 years ago and "seeing if blood pressure is going up or is too low" to make decisions (health care seeking) (Participant 13) and another having saved their laboratory results for the last 10 years "to compare them [and] if off kilter, say[ing] okay, we'll figure out why" (Participant 1). Participants tracked their own trends to understand, inform, and personalize their discussions with providers, as one participant elaborated of using an Excel spreadsheet, to "track everything" noting when they experienced spikes, and graphed trends over time to "be able to join the picture and get right in it and discuss it" with their cardiologist (Participant 23).

In establishing their baseline, participants considered data accuracy as important, at times expressing distrust or concerns about the reliability of their devices and the implications for their decision-making. One participant with a smartwatch questioned the reliability of their smartwatch when it "reported different numbers" compared with that of hospital heart monitors (Participant 30). In deciding about the need for an emergency department (ED) visit "when it [AF] just didn't stop," another participant, who had a smartwatch took "my pulse just to make sure the watch was working" (Participant 3). Between 2 devices' (KardiaMobile and FitBit Sense) reading discrepancies, another participant preferred to rely on their Fitbit as, "my condition looks worse on the Kardia[Mobile] than it does on the Fitbit" (Participant 18). Others had done research into the accuracy of different devices for AF to decide which to use for establishing their own personal baseline. Yet, another participant described the information from their digital technology that varied daily as establishing their own personal reference points:

> *It's like standing on a scale when you go to the doctor's office—it tells you you're always something different than you are at home. But it's relative. I was this much today, and I'm that much tomorrow. So it's all relative to your own personal information.* [Participant 18]

Keeping Out of the Danger Zone

Keeping out of the danger zone was a proactive approach in which patients used technology to "self-diagnose" and validate AF episodes both for making decisions in the short and long terms. Short-term decisions were made to avoid personal risks of triggering an AF episode or respond to symptoms in the moment while long-term decisions focused on preventing complications such as stroke. For example, participants discussed using tracking technology together with their symptoms to monitor for AF triggers to inform behavior change. One participant shared, "I found that my Apple Watch confirmed what my symptoms were already telling me...if it is happening more, I need to change my behaviors i.e. alcohol intake, caffeine intake, amount of social involvement/volunteer commitments,

XSL•FO
RenderX

and personal stress" (Participant 19). This was echoed by another participant who explained that in the short term, they could react and change behaviors in the moment, but also, "I think, in more the longer term the technology gets me better informed what I should or shouldn't be doing" (Participant 27).

Participants' use of technology to keep out of the danger zone was similar whether their AF was symptomatic or asymptomatic. Asymptomatic participants found ongoing tracking invaluable to alert them to AF episodes of which they would otherwise be unaware. Several participants found continuous wearable tracking devices highly valuable, such as "detecting AF during the night when you're asleep" (Participant 14). Even symptomatic participants used continuous wearable technology such as one participant who had experienced a missed AF on a Holter monitor that a continuous wearable detected and was accepted by their provider. Both symptomatic and asymptomatic participants used technology to reduce fear and uncertainty related to their AF risks. For example, technology use reduced an asymptomatic participant's fear of a stroke, "Because with atrial fibrillation, you don't always know when you don't always feel it. I live in fear of having a stroke, and so it just gives me peace of mind" (Participant 4). Symptomatic participants similarly used technology to allay uncertainty, such as one participant with AF symptoms that were "not subtle," who wore a smartwatch daily as a "backup" to allow them to confirm "yeah, this is happening right now, and it started happening then" (Participant 17). One participant who was not considering using a smartwatch until participating in the focus group (despite her brother's recommendation) entertained the value of being able to confirm AF episodes, "You know, because sometimes I think, oh, I wonder if that's AFib, or if it's just because I walked up the big hill" (Participant 20).

Other participants personalized when and what technology they used to meet their individual needs, such as when symptomatic episodes arose, or whether asymptomatic for intermittent monitoring, or just during certain activities (such as exercising). One asymptomatic participant, who never knew when he had an elevated heart rate ("up to 200 or 150") periodically checked his biometrics, with blood pressure or heart rate monitor and KardiaMobile as "without that, I often wouldn't even know I'm having a problem" (Participant 9). Another participant who used a chest strap heart monitor only when exercising also had a KardiaMobile, blood pressure monitor, and oximeter, and limited using these only if symptomatic, admitting, "Most of the time it's through a self-assessment of how I feel during the day and night...If I don't have those symptoms, I don't do anything" (Participant 28). Three participants described using a heart rate strap with their smartwatch to accurately monitor their heart rate response to exercise and use the data to pace themselves to avoid "stress on my heart" (Participant 28), as one participant described:

> *If I'm out for a long run, and I suddenly find that my heart rate, because of the strap, has me over 140 or 150, which isn't that unusual, what I'll do, only because I don't want to red line things, I'll just back off. And I'll slow down, so it allows me to manage my exercise routines as well as my heart health care if it gets into sort of high-risk situations.* [Participant 27]

Other participants used technology for their personal medication management. Participants described using technology to promote medication compliance, such as setting a smartwatch "alarm to remind [me] to take [my] medication" (Participant 18). One participant recalled a specific incident of how technology provided data that allayed their risk-related concerns and gave reassurance as to the effectiveness of their medication management self-care:

> *When my resting heart rate was 104, I was quite concerned about that high level of resting heart rate, but subsequently that has gradually gone down to 78 now. So that helps me detect, you know that I'm taking the right medications and know the medications are working. So that's my feedback mechanism to help me realize that, you know, I'm not in the danger zone.* [Participant 30]

Watchful Waiting

Watchful waiting was a prevalent approach participants used specifically during episodes of AF in making decisions whether or not to seek health care. Watching and waiting was a very active decision-making approach in which participants engaged in 1 or more of the following over time that varied from a few hours to days: watching their digital device and biometric monitor data closely, observing changes in values relative to normal range over time, listening to their symptoms (if they were symptomatic), considering their larger context (eg, what they were doing and what might be triggering), and putting the data together over time to decide whether to seek care. A number of participants were exuberant about the lifesaving role of their tracker that was reinforced by some of their providers.

Technology made a significant difference for participants in their decision-making about seeking ED care. Several participants described watching and waiting using their technology data only, and/or coupled with various sources of data, they processed to inform their decisions. Some sought ED care earlier, later, or not at all, the time frame often varying according to the types and number of data sources they used. Those who sought emergency care earlier often relied exclusively on their immediate device readings, such as one participant who went to the ED within 2 hours of KardiaMobile confirming AF "after one hour of it going on" but which had stopped by ED arrival (Participant 12). Using data from 2 devices (smartwatch and heart rate monitor) to confirm prolonged AF (4 hours), together with considering their activity and symptoms (vague or very mild), another participant decided "pretty quickly that I've got to head into a clinic, or somebody's got to see me pretty fast, and you're off to the emergency room" (Participant 27).

In contrast were some participants who watched and waited over several days and varying activities, during which they monitored their subjective symptoms and objective device data before deciding to seek care. For example, one participant used night-time alerts during sleep showing a high heart rate (180) and a few days later experienced difficulty keeping up during activity that prompted a doctor visit, "I went for a little bike riding [with grand-daughter] and I was having trouble keeping

XSL•FO
RenderX

up with her. And I'm out of breath. Something's wrong here, plus the alert. So I went to the doctor" (Participant 18).

Some participants described avoiding unnecessary ED visits because of the objective reassurance their tracker data gave them despite subjective symptom severity,

> *I would sort of look at it [smartwatch], and my symptoms are really bad, and I'd watch, and I'd think, 'Okay, well, I'm going to give this another few hours.' But I never went to emerge [ED] ever because the watch reassured me that, you know, I was okay hahaha. So I never got frightened that way. I think if I didn't have the watch, I might have been more nervous. I might have been maybe seeking more emergency care.* [Participant 19]

Seeking Decision-Making Support

Despite settling into patterns of independent daily decision-making, participants recognized the need to supplement these in some situations related to their AF self-care often looking to others for support. However, there was considerable variation in participants' need for support that reflected either their comfort or cautiousness in using the data collected from their different technologies—tracking devices, laboratory work portal, and internet—for self-care decision-making. One participant captured this variability, contrasting his wife's "insisten[ce] looking at her results as soon as they come up [and] panics over every number that's outside the normal range, [while] I refuse to go and look at my results. I wait for the doctor" (Participant 3).

Participants expressed considerable awareness of the limitations of relying solely on their technology data for self-care decisions and described situations warranting deferral to their providers. Data misinterpretation was their chief concern whether from lack of context, knowledge, or potential for "life and death" consequences. Participants regularly used online information sources including medical websites (Mayo Clinic) and ChatGPT (Participant 13) for gaining knowledge or answers to their questions about their personal health situation, such as implications of comorbidities (autoimmune diseases) for their AF (Participant 11). Yet, they exercised caution and wanted to check it out with their physicians despite some experiencing decisional pressure from "a lot of people, friends, and relatives—Oh do this, do that, I have done that" (Participant 13). Understanding and interpreting laboratory results were highly variable with some participants comfortable because of previous background (eg, medical field) and others uncomfortable. Participant 4 expressed their discomfort with interpreting laboratory results, comparing the difference between using "trial-and-error" for YouTube music and "dealing with your personal circumstance" but continued tracking patterns and trends after discussing with their general practitioner, blending provider support with independent self-care decision-making. Other areas for which participants wanted supplemental information related both to optimizing their technology data interpretation, such as "all the information I get from my [smart]watch (e.g., nighttime bradycardia)" (Participant 2), and translating their online searches into personal application for their decision-making. One participant, who wanted help with the difficult decision of "when it's time to go to emerge [ED]" wanted more help with data interpretation and suggested use of "Artificial Intelligence, that 'would [be] require[d] to do the work of doctors'" (Participant 17).

Finding the Place for Technology in Normalizing Daily Life

Participants' desire to normalize life with AF was a consistent refrain in their accounts, but the role of technology in doing so varied. For many, it facilitated their feeling of normalcy, while for others it detracted.

Normalizing Life With Technology Use

Several participants found that technology use helped normalize life. Some found that technology gave peace of mind and reassurance, allowing them to get on with life. According to one participant, "I love technology. I almost want to use it so I can get instant information, put it aside, and get on with my life" (Participant 23). Device ECGs showing that they were in normal sinus rhythm were highly reassuring for participants and contributed significantly to normalizing their lives. One participant experienced this reassurance through the habitual use of technology to take smartwatch readings every morning,

> *Most mornings I wake up, and I feel like I'm in AFib. So it's almost more of a method of sort of proving to myself that I'm not in AFib. So I'll just take a reading as reassurance "No, you're fine get on with the day."* [Participant 26]

Technology and specifically the use of home heart rate or rhythm tracking facilitated participants' decision-making confidence in responding to their AF symptoms that helped normalize life and reduce anxiety. Some also talked about needing to track less as they became more attuned to their symptoms, as one participant elaborated:

> *I only use it if I sense it goes wrong. I've had, over the years, I really get this, I can really sense my heart. I know if it's starting to feel like a bouncing tennis ball, or if it starts to do something, I'll get the KardiaMobile out, and then I'll just run a, you know, a few strips.* [Participant 23]

Normalizing Life With Limited Technology Use

Limiting technology use helped normalize life for other participants. They decided that *not* dwelling on AF through the overuse of technology contributed to their sense of normality and reduced stress. Participant 20, who initially expressed discomfort using a personal device because, "if you're obsessing over your AFib, you're stressing yourself out" which she considered a possible AF trigger, normalized life by just sticking with her doctor's prescribed testing or monitoring. Reinforcing this sentiment, another participant agreed, saying "don't fuss over AFib. Just [live] life as normal as possible" (Participant 3). Another participant who admitted that they could "get to be a bit obsessive about data...but did not have a context in which to analyze it" normalized life by letting go of the data obsession and now monitors only their blood pressure periodically at the recommendation of their general practitioner, and "I just look at spikes, otherwise I just get on with life" (Participant 4). A

XSL•FO
RenderX

participant who had "faithfully" monitored blood pressure for 2 years after diagnosis until the COVID-19 pandemic and had not resumed, "I've just kind of, I know when I'm in AF, I can feel it and I change what I do, kind of calm myself down a bit, and it seems to help it stop," is no longer monitoring except occasionally checks it out as symptoms arise, "otherwise I'm just living my life the way I used to. Not worrying about it" (Participant 5).

Despite settling into a rhythm with technology as a tool to help them get on with their lives, for some there was an underlying uncertainty about how technology could render them AF-free which for them would return them to a sense of normalcy. Ultimately, one pondered the process of coming to the place where technology fits in self-care decision-making, amid the uncertainty of their AF trajectory,

> *I find all of this is allowing me to monitor, and I guess, manage, on a day-to-day basis. But in terms of what it means in the greater picture, I really don't know. That to me is kind of the ultimate end. Some sort of resolution to this.* [Participant 27]

Discussion

Principal Findings

This study provides novel qualitative insights into the experiences of older adults with AF using digital technology to make decisions about their daily self-care. The extent to which patients made use of technology in their self-care decision-making was highly personalized as they considered their context, symptoms, and activities. This process was characterized by several interrelated themes, beginning with either patient-initiated or provider-recommended use, and developing into patterns of digitally supported decision-making, and finally finding the place for technology within their life context.

Findings reflected that technology facilitated participants' independent or naturalistic decision-making about their daily AF self-care, which is largely performed by patients outside of health care settings [6]. Similar to another study of patients with AF, whose use of heart-tracking technologies gave confirmation and reassurance about their bodily symptoms, patients in this study often gained self-care decision-making confidence and reassurance from their technology-collected objective data [21]. This resonates with and extends the naturalistic decision-making literature, including the illustration of the Recognition-Primed Decision Model in self-care decision-making by Riegel et al [11] and the process by Daley et al [12] regarding monitoring, interpreting, and acting among patients with heart failure by showcasing the growing centrality of technology in decision-making for both habitual and irregular users. AF's unpredictability and uncertain course make it a highly challenging condition to manage, often leaving patients feeling left to trial and error [3,4]. Our in-depth look at patterns of daily self-care decision-making highlighted ways older adults with AF used technology to improve their situation awareness (eg, reduce symptom uncertainty or ambiguity), supporting sophisticated naturalistic decision-making.

Our participants recognized the value of continuous monitoring with wearable devices and smartwatches to make actionable decisions to avoid triggers, reduce risks, modify lifestyle behaviors and daily activities, and guide health care seeking. These findings contrast with those from a study on technology use among patients with rheumatoid arthritis, who questioned the value of pain monitoring and tracking if it was not going to lead to any change in their management [31]. This divergence suggests that self-care decision-making using technology is highly disease-specific [22]. A comparative qualitative study of the self-care decision-making between patients with type 2 diabetes, HIV/AIDS, and multiple sclerosis found that there were common yet disease-specific self-care decisions participants made because each disease has unique attributes that influence the meaning and interpretation of specific self-care decisions [32].

Participants in this study made judicious use of health care services and tried to avoid unnecessary ED visits, as they used watchful waiting strategies in making their decisions. Although concerns have been raised about the potential for consumer device use to prompt greater health care use in patients with AF [33,34], our participants detailed decision-making experiences of knowing *better* when to seek health care, reinforcing notions that technology holds great potential for win-win solutions that both support patients and lessen burden on health systems [35]. Accounts from individuals who used self-tracking technology pre–cardiac diagnosis also resonate with our findings as they collected data over time, compared with their norm, and continued to monitor before deciding to seek care that ultimately led to diagnoses [36]. Nevertheless, the use of consumer devices for widespread AF screening is highly contested and carries potential for medical overuse among undiagnosed people who are unlikely to gain clinical benefit [37].

Although participants in this study emphasized their independent use of technology for making their self-care decisions, they also collected, tracked, and trended their data to share and/or discuss with their providers and deferred to them to meet their need for support interpreting data (eg, laboratory values) in making decisions. In previous intervention research, participants who provided a smartphone app for AF management expressed disappointment to find that there was no mechanism for the data collected to be shared directly with their providers [7]. Although our participants were not directly provided devices, they too were often seeking decision-making supports to process their personal data. Several patients in this study also expressed concerns about anxiety and becoming overly preoccupied with tracking. Indeed, the potential for consumer heart-tracking devices to provoke anxiety in patients has been highlighted in other work [38]. In an intervention study of 14 patients with AF who were provided several monitoring devices, many found the measurements motivational, but 2 patients expressed concern about overmonitoring [19].

Findings revealed variability in the way patients normalized their use of technology for self-care decision-making. For many patients, technology supported informed decision-making that provided them reassurance, allowing for a return to a sense of normalcy, whereas for others, normalizing life was not becoming preoccupied with the use of technology that kept them fixated

XSL•FO
RenderX

on their AF. Similar to the process of finding normality in living with chronic disease [39], patients integrated technology within that process as they adapted and constantly defined and redefined normalcy. Overall, technology use for achieving normalcy had to align with patients' personal preferences, reflected in the growing body of literature on the need for technology to be personalized according to the patient and the purpose [40,41]. Importantly, consideration must be given to contextual factors (eg, cost, preferences, and digital literacy) that influence technology use to avoid reinforcing unequal access to health care services and disparities in socioeconomic status for marginalized older adults with AF [42].

These findings have implications for providers, patients, researchers, and policymakers. While a care goal of providers is to encourage patient self-care and active involvement in their care, they are often limited in supporting patients' naturalistic decision-making, for example, in implementing and sustaining lifestyle changes [43]. Given this study's elucidation of perceptions of patients with AF regarding the value of technology to support their self-care decision-making, providers might choose to encourage the use of these tools. Supporting them, however, may require additional provider education and training [42]. The 4 decision-making approaches observed in this work hold potential to inform the development of patient-facing tools to better support patients with AF in naturalistic decision-making by building their capacity to interpret and process the growing volumes of data they obtain with technology and to reduce clinician burden. Canadian Cardiovascular Guidelines for AF [44] include recommendations related to self-care and lifestyle but have not yet incorporated the role of consumer devices in these recommendations. Policy guidance is needed on best practices to inform patients and providers on the most appropriate technologies, personalizing use to patient needs, and a balanced incorporation of consumer data in decision-making.

Limitations

Focus group participants were well educated, were fairly active users of technology in general, and had higher incomes, which may limit the generalizability of our findings to others. The cost of consumer devices, for example, may be a significant barrier in other samples that include greater representation of patients with lower socioeconomic backgrounds. Virtual focus groups allowed participants to join from wherever it was most convenient for them and may have attracted a biased sample, although 3 participants did join using the telephone option. Although we reached participants with a broad range of technology use, including those who described themselves as low technology users, the sample may not adequately represent marginalized or genuinely low-technology populations, given all participants used at least 1 device (computer, smartphone, or tablet), and technology use was defined broadly (eg, "use of the internet") during recruitment. Nevertheless, the proportion using heart trackers mirrored that of a national survey [20].

Conclusions

This study contributes the first in-depth qualitative insights into the use of technology in naturalistic decision-making among older adults with AF. Our patient participants used technology to personalize everyday AF self-care, outlining intricate decision-making processes. These findings have implications for supporting decision-making tailored to individual needs and context while considering the broad range of technology use among patients with AF. The findings can also form the basis for continuing work in this area to further investigate the role technology plays in patient decision-making and quality of life, as well as health care use, costs, and provider burden.

Acknowledgments

The authors would like to thank Kai Derksen for assistance facilitating focus groups. The authors declare that this work was written entirely by the author team without the assistance of generative artificial intelligence tools.

Funding

This research was supported by a Canadian Institute of Health Research Planning and Dissemination grant (GR031722, 2024), a Michael Smith Health Research BC Convening and Collaborating (C2) Award (GR033520, 2024), and the UBC Okanagan's Work Study program (project 833733). The funders were not involved in the design, execution, analysis and interpretation, writing, or decision to submit this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the conceptualization (formulation or evolution) and methodology. Additionally, KLR participated in formal analysis, funding acquisition, investigation, supervision, validation, writing – original draft, and writing – review and editing. CLS participated in data curation, project administration, writing – original draft, and writing – review and editing. AL participated in formal analysis, validation, writing – original draft, and writing – review and editing. VW participated in investigation (recruitment) and writing – review and editing. PSL participated in writing – review and editing. WC, MP, LM, and JGA contributed to resources and writing – review and editing. All authors have read and agreed to the published version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Illustration of main themes derived inductively through initial codes and subcategories identified.
[DOCX File, 17 KB - cardio_v10i1e93036_app1.docx]

References

1. Riegel B, Jaarsma T, Strömberg A. A middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2012;35(3):194-204. [doi: 10.1097/ANS.0b013e318261b1ba] [Medline: 22739426]
2. Pathak RK, Evans M, Middeldorp ME, et al. Cost-Effectiveness and Clinical Effectiveness of the Risk Factor Management Clinic in Atrial Fibrillation: The CENT study. JACC Clin Electrophysiol 2017 May;3(5):436-447. [doi: 10.1016/j.jacep.2016.12.015] [Medline: 29759599]
3. Stridsman M, Strömberg A, Hendriks J, Walfridsson U. Patients' experiences of living with atrial fibrillation: a mixed methods study. Cardiol Res Pract 2019;2019:6590358. [doi: 10.1155/2019/6590358] [Medline: 31885904]
4. McCabe PJ, Schumacher K, Barnason SA. Living with atrial fibrillation: a qualitative study. J Cardiovasc Nurs 2011;26(4):336-344. [doi: 10.1097/JCN.0b013e31820019b9] [Medline: 21263348]
5. Vest BM, Quigley BM, Lillvis DF, et al. Comparing patient and provider experiences with atrial fibrillation to highlight gaps and opportunities for improving care. J Gen Intern Med 2022 Sep;37(12):3105-3113. [doi: 10.1007/s11606-021-07303-5] [Medline: 34993876]
6. Riegel B, Moser DK, Buck HG, et al. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):e006997. [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
7. Mihas P, Rosman L, Armbruster T, Walker J, Deyo Z, Gehi A. Patient perspectives on performance of a smartphone app for atrial fibrillation self-management. Patient Prefer Adherence 2022;16:2799-2810. [doi: 10.2147/PPA.S366963] [Medline: 36281351]
8. Klein G. Naturalistic decision making. Hum Factors 2008 Jun;50(3):456-460. [doi: 10.1518/001872008X288385] [Medline: 18689053]
9. Klein GA, Calderwood R, Clinton-Cirocco A. Rapid decision making on the fire ground. Proc Hum Factors Soc Annu Meet 1986 Sep;30(6):576-580. [doi: 10.1177/154193128603000616]
10. Dhukaram AV, Baber C. Modelling elderly cardiac patients decision making using cognitive work analysis: identifying requirements for patient decision aids. Int J Med Inform 2015 Jun;84(6):430-443. [doi: 10.1016/j.ijmedinf.2015.01.001] [Medline: 25805169]
11. Riegel B, Dickson VV, Topaz M. Qualitative analysis of naturalistic decision making in adults with chronic heart failure. Nurs Res 2013;62(2):91-98. [doi: 10.1097/NNR.0b013e318276250c] [Medline: 23154659]
12. Daley CN, Cornet VP, Toscos TR, Bolchini DP, Mirro MJ, Holden RJ. Naturalistic decision making in everyday self-care among older adults with heart failure. J Cardiovasc Nurs 2022;37(2):167-176. [doi: 10.1097/JCN.0000000000000778] [Medline: 33369989]
13. Linz D, Gawalko M, Betz K, et al. Atrial fibrillation: epidemiology, screening and digital health. Lancet Reg Health Eur 2024 Feb;37:100786. [doi: 10.1016/j.lanepe.2023.100786] [Medline: 38362546]
14. Lane DA, McMahon N, Gibson J, et al. Mobile health applications for managing atrial fibrillation for healthcare professionals and patients: a systematic review. Europace 2020 Aug 27;22(10):1567-1578. [doi: 10.1093/europace/euaa269] [Medline: 32853369]
15. Wilson RE, Burton L, Marini N, et al. Assessing the impact of atrial fibrillation self-care interventions: a systematic review. Am Heart J Plus 2024 Jul;43:100404. [doi: 10.1016/j.ahjo.2024.100404] [Medline: 38831787]
16. Trivedi R, Shaw T, Sheahen B, Chow CK, Laranjo L. Patient perspectives on conversational artificial intelligence for atrial fibrillation self-management: qualitative analysis. J Med Internet Res 2025 Mar 12;27:e64325. [doi: 10.2196/64325] [Medline: 40073398]
17. Nathania J, Woo BFY, Cher BP, et al. Patient perspectives of the Self-management and Educational Technology tool for Atrial Fibrillation (SETAF): a mixed-methods study in Singapore. PLoS One 2022;17(1):e0262033. [doi: 10.1371/journal.pone.0262033] [Medline: 35061749]
18. van Mourik MJW, Keijsers L, van der Velden RMJ, et al. Patients perspectives on integrating eHealth in regular care pathways for atrial fibrillation: evaluating photoplethysmography for remote self-assessment. Eur J Cardiovasc Nurs 2025 Mar 3;24(2):305-313. [doi: 10.1093/eurjcn/zvae156] [Medline: 39749467]
19. Joensen EDR, Albertsen AE, Spindler H, et al. Patient perspectives on the "future patient" telerehabilitation program for atrial fibrillation: qualitative study. JMIR Cardio 2025 Aug 19;9:e68663. [doi: 10.2196/68663] [Medline: 40829146]
20. Hills MT. Patient perspective: digital tools give afib patients more control. Cardiovasc Digit Health J 2021 Jun;2(3):192-194. [doi: 10.1016/j.cvdhj.2021.05.001] [Medline: 35265908]

XSL•FO
RenderX

21. Keys R, Marshall P, Stuart G, O'Kane AA. Rethinking lived experience in chronic illness: navigating bodily doubt with consumer technology in atrial fibrillation self-care. 2025 Apr 26 Presented at: Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2025; Yokohama Japan p. 1-19. [doi: 10.1145/3706598.3713326]
22. Riegel B, Jaarsma T, Strömberg A. An update to the middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2026;49(1):14-25. [doi: 10.1097/ANS.0000000000000594] [Medline: 40956127]
23. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. Res Nurs Health 1997 Apr;20(2):169-177. [doi: 10.1002/(SICI)1098-240X(199704)20:2<169::AID-NUR9>3.0.CO;2-I] [Medline: 9100747]
24. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Ageing. World Health Organization. 2023. URL: https://www.who.int/health-topics/ageing#tab=tab_1 [accessed 2026-04-09]
26. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191] [Medline: 847061]
27. Hennink MM, Kaiser BN, Weber MB. What influences saturation? Estimating sample sizes in focus group research. Qual Health Res 2019 Aug;29(10):1483-1496. [doi: 10.1177/1049732318821692] [Medline: 30628545]
28. Tracy SJ. Qualitative quality: eight "Big-Tent" criteria for excellent qualitative research. Qual Inq 2010 Dec;16(10):837-851. [doi: 10.1177/1077800410383121]
29. Richards L, Morse JM. Readme First for a User's Guide to Qualitative Methods: SAGE Publications, Inc; 2013. [doi: 10.4135/9781071909898]
30. NVivo leading qualitative data analysis software (QDAS). Lumivero. 2024. URL: https://lumivero.com/products/nvivo/ [accessed 2024-12-22]
31. Wada M, Wallace JR. Designing technologies for self-care: describing the lived experiences of individuals with rheumatoid arthritis. Hum Factors Healthc 2022 Dec;2:100025. [doi: 10.1016/j.hfh.2022.100025]
32. Paterson B, Thorne S, Russell C. Disease-specific influences on meaning and significance in self-care decision-making in chronic illness. Can J Nurs Res 2002 Oct;34(3):61-74 [FREE Full text] [Medline: 12425011]
33. Rosman L, Lampert R, Zhuo S, et al. Wearable devices, health care use, and psychological well-being in patients with atrial fibrillation. J Am Heart Assoc 2024 Aug 6;13(15):e033750. [doi: 10.1161/JAHA.123.033750] [Medline: 39011944]
34. Wang L, Nielsen K, Goldberg J, et al. Association of wearable device use with pulse rate and health care use in adults with atrial fibrillation. JAMA Netw Open 2021 May 3;4(5):e215821. [doi: 10.1001/jamanetworkopen.2021.5821] [Medline: 34042996]
35. Aljuaid M, Marashly Q, AlDanaf J, et al. Smartphone ECG monitoring system helps lower emergency room and clinic visits in post-atrial fibrillation ablation patients. Clin Med Insights Cardiol 2020;14:1179546820901508. [doi: 10.1177/1179546820901508] [Medline: 32009826]
36. Keys R, Marshall P, Stuart G, O'Kane AA. "I think it saved me. I think it saved my heart": the complex journey from self-tracking with wearables to diagnosis. 2024 May 11 Presented at: Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu HI USA p. 1-15. [doi: 10.1145/3613904.3642701]
37. Shih P, Prokopovich K, Degeling C, Street J, Carter SM. Direct-to-consumer detection of atrial fibrillation in a smartwatch electrocardiogram: Medical overuse, medicalisation and the experience of consumers. Soc Sci Med 2022 Jun;303:114954. [doi: 10.1016/j.socscimed.2022.114954] [Medline: 35569232]
38. Rosman L, Gehi A, Lampert R. When smartwatches contribute to health anxiety in patients with atrial fibrillation. Cardiovasc Digit Health J 2020;1(1):9-10. [doi: 10.1016/j.cvdhj.2020.06.004] [Medline: 34386784]
39. Robinson CA. Managing life with a chronic condition: the story of normalization. Qual Health Res 1993 Feb;3(1):6-28. [doi: 10.1177/104973239300300102] [Medline: 8457793]
40. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171. [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
41. Varma N, Han JK, Passman R, et al. Promises and perils of consumer mobile technologies in cardiovascular care: JACC scientific statement. J Am Coll Cardiol 2024 Feb 6;83(5):611-631. [doi: 10.1016/j.jacc.2023.11.024] [Medline: 38296406]
42. Qiu Y, Asselin CY, Kallas D, et al. The emerging role of digital health technology in cardiovascular care. Can J Cardiol 2021 Jul;37(7):939-942. [doi: 10.1016/j.cjca.2021.04.016] [Medline: 34600793]
43. Lee G, Baker E, Heidbuchel H, et al. Patients' experiences of atrial fibrillation and comorbidity management in clinical practice: a Pan-European qualitative descriptive interview study. BMJ Open 2025 Jun 16;15(6):e094839. [doi: 10.1136/bmjopen-2024-094839] [Medline: 40523789]
44. Andrade JG, Aguilar M, Atzema C, et al. The 2020 Canadian Cardiovascular Society/Canadian Heart Rhythm Society comprehensive guidelines for the management of atrial fibrillation. Can J Cardiol 2020 Dec;36(12):1847-1948. [doi: 10.1016/j.cjca.2020.09.001] [Medline: 33191198]

XSL•FO
RenderX

Abbreviations

AF: atrial fibrillation
COREQ: Consolidated Criteria for Reporting Qualitative Research
ECG: electrocardiogram
ED: emergency department

Edited by A Coristine; submitted 09.Feb.2026; peer-reviewed by D Lai, G Lee; revised version received 01.May.2026; accepted 04.May.2026; published 11.Jun.2026.

Please cite as:
Rush KL, Seaton CL, Lukey A, Wang V, Loewen PS, Chiu W, Patapoff M, Moroz L, Andrade JG
Experiences of Patients With Atrial Fibrillation Using Technology to Personalize Self-Care Decision-Making: Interpretive Description Study
JMIR Cardio 2026;10:e93036
URL: https://cardio.jmir.org/2026/1/e93036
doi:10.2196/93036

© Kathy L Rush, Cherisse L Seaton, Alexandra Lukey, Victor Wang, Peter S Loewen, Wynne Chiu, Megan Patapoff, Lana Moroz, Jason G. Andrade. Originally published in JMIR Cardio (https://cardio.jmir.org), 11.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Positive Psychological Well-Being and Determinants of Social Robot Acceptability Among Patients With Heart Failure: Cross-Sectional Questionnaire Study

Lisa-Marie Maukel[1], PhD; Karen Bouchard[1,2], PhD; Jess G Fiedorowicz[2,3], MD, PhD; Kerstin Dautenhahn[4], Dr.rer.nat; Moojan Ghafurian[4], PhD; Thais Coutinho[5], MD, FRCPC; Caroline McGuinty[2,6], MD, FRCPC; Peter P Liu[2,6], MD, FRCPC; Heather Tulloch[1,2], PhD, C.Psych

[1]Division of Prevention and Rehabilitation, University of Ottawa Heart Institute, 40 Ruskin Street, l, Ottawa, ON, Canada

[2]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[3]Department of Mental Health, The Ottawa Hospital, Ottawa, ON, Canada

[4]Department Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[5]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, United States

[6]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada

Corresponding Author:
Heather Tulloch, PhD, C.Psych
Division of Prevention and Rehabilitation, University of Ottawa Heart Institute, 40 Ruskin Street, l, Ottawa, ON, Canada

Abstract

Background: Social robots (SRs) are innovative tools in health care, offering both medical and psychological support for patients with heart failure (HF). For successful implementation, patient acceptability of SRs is crucial. Living in urban areas and having a lower comorbidity burden have been linked to higher acceptability; however, the role of psychological factors remains underexplored.

Objective: This study aimed to examine the associations between negative (eg, depression and anxiety) and positive (eg, optimism) psychological factors and personality traits (eg, openness and extraversion) with SR acceptability in patients with HF.

Methods: Patients with HF watched brief videos about SRs and completed validated measures of depressive symptoms (Patient Health Questionnaire-9), anxiety symptoms (Generalized Anxiety Disorder-7), positive psychological well-being (Brief Inventory of Thriving), and personality traits (Ten-Item Personality Inventory). Medical information was extracted from patients' records. SR acceptability was assessed using the Unified Theory of Acceptance and Use of Technology (UTAUT). Pearson correlations and multiple linear regression, adjusted for age, sex, smart technology experience, urbanicity, and comorbidities, were conducted.

Results: Of the 101 patients (women: n=36, 35.6%, mean age 68, SD 10 y), 23% (23/101) scored in the clinical range for depression, and 17% (17/101) scored in the clinical range for anxiety. Well-being scores were moderate, and conscientiousness and agreeableness were the most common. UTAUT behavioral intention was moderate; 69% (67/97) of participants were likely to use an SR if available. Well-being scores correlated positively with SR acceptability in 4 of 5 UTAUT subscales, whereas no significant bivariate associations were observed for psychological distress or personality traits. In the multiple regression models, higher Brief Inventory of Thriving scores were associated with increased SR acceptability, including UTAUT facilitating conditions (B=0.17; P=.01) and behavioral intention (B=0.17; P=.04), independent of depressive and anxiety symptoms.

Conclusions: Psychological well-being is associated with determinants of SR acceptability in patients with HF, while psychological distress and personality traits are not associated with these determinants. These patient-level factors ought to be examined more closely before SR implementation.

(JMIR Cardio 2026;10:e83163) doi:10.2196/83163

KEYWORDS

heart failure; social robots; robotics; technology acceptability; positive psychological well-being; assistive technology; mental health

Introduction

Heart failure (HF) is a growing global health concern, affecting approximately 56.2 million people worldwide [1]. The condition is characterized by fluctuating functional impairment, frequent exacerbations, and high hospitalization rates, with nearly 50% of patients with HF being readmitted after an initial hospital stay [2]. These recurrent hospitalizations lead to a significant

XSL•FO
RenderX

decline in quality of life and contribute to a 5-year mortality rate of 50% [3,4]. Moreover, HF imposes substantial economic burdens, with annual health care costs in the United States reaching US $30.7 billion [5]. The unpredictable trajectory, severe symptom burden of HF [4], and the high clinical demand [6] highlight the need for innovative solutions to improve patient support and disease management, and reduce resource burden.

One emerging technological advancement is the use of social robots (SRs) [7]. These artificial intelligence–powered agents, designed to mimic human or animal characteristics, engage users through verbal, nonverbal, and affective interactions [8]. SRs can assist in a variety of areas, including health monitoring (eg, blood pressure tracking), daily self-care tasks (eg, medication adherence), rehabilitation, and providing companionship [8-10]. By simulating health care provider interactions and facilitating home-based care [11], SRs align with the increasing demand for strategies that reduce hospital readmissions and improve patient satisfaction, particularly for those with limited access to health care services [12].

Research in older adults has demonstrated that SRs can positively impact physiological markers, such as pulse rate, cortisol levels, blood pressure, and oxygenation [13]. Additionally, SRs have been shown to promote exercise and medication adherence, reduce unhealthy snacking, and support weight loss [14-18]. Among those with dementia, SRs have effectively addressed psychosocial needs by alleviating loneliness and depressive symptoms [19-21], factors particularly relevant for patients with HF [22].

Despite these promising findings, the application of SRs in cardiac care remains limited. Before implementing novel technology, assessing acceptability within the target population is crucial [11,23,24]. In contrast to acceptance, which refers to the actual adoption or sustained use of a technology, acceptability reflects an individual's behavioral intention to use a technology, influenced by factors such as performance expectations, usability, privacy, and patient characteristics [25-27]. So far, only one small study (n=3) examined SR use in patients with cardiovascular conditions, suggesting that SRs may improve exercise adherence during cardiac rehabilitation compared with usual care [28]. An accompanied survey of 28 patients with cardiac conditions revealed that 75% responded positively to SRs following direct interaction [29]. Our prior research involving HF health care providers, including cardiologists and cardiac nurses, indicated a general openness to using SRs for data collection and patient prompts [30]. Similarly, our study including patients with HF found that 69% would be willing to use an SR if available, with higher acceptability observed among those living in urban areas and experiencing a lower symptom burden [31]. These findings offer initial insights into potential determinants of SR acceptability in the HF population.

Psychological factors likely play a key role in shaping SR acceptability. Depressive symptoms are associated with reduced approach motivation and behavioral withdrawal, whereas anxiety increases avoidance and threat sensitivity, particularly in contexts involving uncertainty [32]. Within this framework, depressive and anxious symptoms may act as barriers to SR acceptability. However, empirical findings are inconsistent: some studies report higher SR acceptability with negative affect [33], while another study found no such connection with depressive mood [34]. Although technology-specific anxiety consistently predicts lower SR acceptability [24,27,35], the influence of generalized anxiety remains unclear, both in healthy individuals and in clinical populations. One experimental study found that fear decreased the likelihood of interacting with SRs in a student sample [36], but its relevance to HF populations remains unclear. Thus, further investigation into the relationship between anxiety and depression with SR acceptability is warranted.

In contrast, positive psychological factors, such as positive psychological well-being (PPWB), activate approach-oriented systems that promote exploration, goal-directed behavior, and openness to novelty [32,37]. In the context of SRs, positive emotions, such as enjoyment and excitement, may increase acceptability of SRs [36,38], potentially by reducing perceived effort and fostering engagement [39].

PPWB, including optimism, purpose, resilience, gratitude, positive affect, and happiness, is increasingly recognized as a critical factor in HF care [40-42]. The 2021 American Heart Association scientific statement on *psychological health, well-being, and the mind-heart-body connection* underscores the association of PPWB with reduced cardiovascular disease risk and lower mortality rates [43]. Importantly, PPWB represents a construct distinct from the mere absence of depression or negative affect and is independently associated with health outcomes [44,45]. However, its relevance to SR acceptability in cardiac populations has not yet been examined.

Beyond affective states, personality traits may also influence SR acceptability. The Five-Factor Model (Big Five) is the most widely recognized framework for describing individual differences in personality [46], comprising neuroticism, extraversion, agreeableness, conscientiousness, and openness to experience [47]. Within an approach-avoidance framework, these traits may shape how individuals evaluate and engage with SRs. Extraversion, characterized by social engagement and reward sensitivity, aligns with approach tendencies and may be associated with more positive attitudes toward SRs. In contrast, neuroticism, marked by threat sensitivity and negative affect, aligns with avoidance tendencies and may reduce acceptability [48]. Openness to experience may enhance receptivity to novel technologies, while conscientiousness and agreeableness may foster engagement through perceived usefulness and trust in the SR as a supportive tool.

These traits have already been shown to be essential for understanding health behaviors and cardiovascular outcomes [49]. Specifically, agreeableness and conscientiousness have been linked to better adherence to treatment regimens [50], which could also apply to SR interventions delivered by health care providers. Although traits such as agreeableness, extraversion, and openness have been found to influence SR acceptability in general populations [51,52], their impact on patients with HF remains unknown.

Understanding the relationship between these psychological factors and SR acceptability in patients with HF is crucial for

XSL•FO
RenderX

developing SR interventions that cater to patients' specific needs and preferences, potentially improving engagement and overall treatment outcomes. As such, this study aimed to examine the relationships between negative (depression and anxiety) and positive (PPWB) psychological factors, personality traits, and SR acceptability in patients with HF.

Methods

Study Population and Recruitment

Patients with HF were recruited from the University of Ottawa Heart Institute, a large tertiary and quaternary cardiac care center serving a population of 1.4 million in the Champlain Local Health Integration Network, Ontario, Canada. Outpatient participants were enrolled between September 2022 and December 2023. Eligible participants had a diagnosis of either HF with preserved ejection fraction (left ventricular ejection fraction [LVEF] ≥50%) or HF with reduced ejection fraction (HFrEF; LVEF <40%) and were symptomatic, as defined by New York Heart Association (NYHA) functional class II-IV, indicating at least mild limitations in physical activity [53]. Additional inclusion criteria included being over 18 years of age, proficient in reading either English or French, and capable of providing informed consent. Exclusion criteria included individuals with unstable psychiatric conditions (eg, suicidal ideation or unmanaged psychotic disorders) or severe cognitive impairment requiring substantial assistance with daily activities.

Participants were recruited through clinician referrals and a review of medical records. Potentially eligible participants were contacted either via phone or in person at the clinic. A research coordinator confirmed eligibility and obtained informed consent. To reduce selection bias, we attempted to approach all eligible patients during the study period. Participants had the option to complete the survey using a secure web-based platform, REDCap (Research Electronic Data Capture) [54], or on paper, either on-site or at home. Before completing the study questionnaires, all patients were shown 4 brief informational manufacturer videos of SR models (Mabu, Catalia Health Inc.; PARO, Paro Therapeutic Robot; Elli-Q, Intuition Robotics; and Pepper, SoftBank Robotics), following established methods to introduce participants to new technologies [55]. The videos demonstrated SR functions such as scheduling appointments, reminding patients about daily exercises and medications, and offering companionship and social connection. Full descriptions and transcripts of all videos are provided in Supplement 1 in Multimedia Appendix 1. For participants without access to the necessary technology, the videos were presented on a tablet during their clinic visit.

Ethical Considerations

This study received approval from the Ottawa Health Science Network Research Ethics Board (Protocol ID: 20210165 - 01H). Written informed consent was obtained from all participants. All study data were deidentified prior to analysis to protect participant privacy. Only study IDs were used, and no personal identifiers were linked to the dataset. Participants received no compensation.

Measurements

Demographic and Medical Information

Participants completed a questionnaire of demographic characteristics (eg, sex, urbanicity, and ethnicity) and medical conditions. These data were verified by their hospital records. Urbanicity was dichotomized to compare urban versus nonurban residence (1=urban; and 0=suburban, mixed, or rural). The Charlson Comorbidity Index (CCI), a weighted tool that quantifies comorbid conditions by accounting for both their number and severity, was calculated to assess the comorbidity burden. Scores range from 0 to 37, with higher scores indicating a greater burden of comorbid diseases [56].

General Attitudes Toward Technology and SRs

On the basis of previous research on SR acceptability in older adults [23], participants completed an 11-item questionnaire assessing their general attitudes toward and experiences with technology and SRs. Example items included "Have you interacted with a social robot in the past?" (yes or no) and "I would use a social robot if it were to become available to me" (agree, somewhat agree, somewhat disagree, or disagree) dichotomized to "Would likely use SR" (yes or no). To capture prior experience with smart technology, we calculated a sum score based on the question "Please select what type of smart devices do you currently have?" Options included smartphone, wearable devices, voice assistant agents, and robotic vacuum cleaners, resulting in a possible range of 0 to 12 smart devices in use. In addition, patients ranked the perceived usefulness of various SR capabilities and features using a 25-item questionnaire [10,23], for providing in-home care. Each item was rated on a 5-point Likert scale, ranging from 1 (extremely useless) to 5 (extremely useful). Participants evaluated capabilities such as management and communication (eg, scheduling appointments), monitoring (eg, tracking blood pressure), medical procedures and treatment (eg, administering oral medications), emotional support (eg, providing companionship), and daily living assistance (eg, assisting with eating). This questionnaire is provided in Supplement 2 in Multimedia Appendix 1.

Depressive Symptoms

Depressive symptoms were evaluated using the Patient Health Questionnaire-9 (PHQ-9) [57,58], a well-established instrument frequently applied with cardiovascular populations [59]. Participants rated 9 items on a 4-point Likert scale, ranging from 0 (not at all) to 3 (nearly every day), generating a total score range from 0 to 27, with higher scores reflecting higher symptom severity. A score of 10 or higher was used as a cutoff for moderate depressive symptoms [57,60]. The tool exhibits strong psychometric properties, including good internal consistency (Cronbach α=0.89) and robust validity [57]. The PHQ-9 demonstrated good internal consistency in this sample (Cronbach α=0.86).

Generalized Anxiety

Anxiety symptoms were assessed using the Generalized Anxiety Disorder-7 (GAD-7) scale [61]. Participants rated 7 items on a 4-point Likert scale, ranging from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 21. Higher scores

XSL•FO
RenderX

indicate greater anxiety symptom severity. A score of 10 or higher is considered the optimal threshold for screening generalized anxiety. In a cardiac sample, the GAD-7 demonstrated strong reliability, with a Cronbach α of 0.89 and a composite reliability index of 0.90. It also exhibited robust validity across procedural, criterion, construct, and factorial measures [62]. The GAD-7 demonstrated excellent internal consistency in this sample (Cronbach α=0.93).

Positive Psychological Well-Being

PPWB was assessed using the Brief Inventory of Thriving (BIT) [63], a 10-item scale measuring positive functioning across multiple well-being domains, including life satisfaction, positive emotions, supportive relationships, engagement, purpose, mastery, and optimism. Each item was scored on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree), with higher scores representing greater well-being. Total scores ranged from 10 to 50. Sample items included "There are people who appreciate me as a person" and "What I do in life is valuable and worthwhile." The BIT has been shown to have strong psychometric properties, including retest reliability and convergent validity [63]. The BIT demonstrated excellent internal consistency in this HF sample (Cronbach α=0.92).

Personality Style

The Ten-Item Personality Inventory (TIPI) [64] was used to assess the 5-factor model of personality: extraversion, agreeableness, conscientiousness, emotional stability (neuroticism), and openness to experience. Two items representing each trait are rated on a 7-point Likert scale, ranging from 1 (disagree strongly) to 7 (agree strongly). Higher scores reflected stronger tendencies toward the respective traits. The TIPI demonstrated acceptable test-retest reliability, structural validity, and convergent validity [65]. The TIPI has been applied before in cardiac samples [66]. In this HF sample, Cronbach α was 0.40 for extraversion, 0.33 for agreeableness, 0.33 for conscientiousness, 0.66 for emotional stability, and 0.49 for openness. These modest values are consistent with expectations for 2-item TIPI subscales [64].

Acceptability of SRs

To assess patients' acceptability of SRs, we used the 19-item Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire [67,68] adapted for SRs. Each item was rated on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). The UTAUT evaluates 5 key factors that influence the behavioral intention to use SRs, with subscale scores ranging from 4 to 28. Higher scores on each subscale indicate greater acceptability and a higher likelihood of SR usage. The factors assessed include performance expectancy (eg, "I would find a social robot useful in my everyday life"), effort expectancy (eg, "I would find the social robot easy to use"), facilitating conditions (eg, "I would have the resources necessary to use a social robot"), and social influence (eg, "People who are important to me think that I should use a social robot"). Behavioral intention to use an SR, with a score range of 3 to 21, was measured with items such as "If it was available to me, I intend to use a social robot." The UTAUT has been used in health care settings involving SRs [69], and its construct validity has been established through confirmatory factor analysis [70]. In this sample, the UTAUT subscales demonstrated excellent internal consistency for performance expectancy (Cronbach α=0.97), effort expectancy (Cronbach α=0.96), social influence (Cronbach α=0.87), and behavioral intention (Cronbach α=0.91), and questionable internal consistency for facilitating conditions (Cronbach α=0.60).

Statistical Analysis

Analyses were conducted in IBM SPSS (version 29). As missing data were minimal (0% - 8% per variable) and assumed missing at random (Little's Missing Completely At Random test: χ^2_{1387}=1415.3; *P*=.29), we used the semiparametric multiple imputation (m=5) procedure [71] to address missing values in the psychological questionnaires. Descriptive statistics were calculated for demographic, clinical, and psychological variables. Continuous variables were reported as means (SDs), and categorical variables were reported as frequencies and percentages. Pearson correlations examined bivariate variable relationships, following Hemphill's guidelines for correlation interpretation: <0.20 (weak), 0.20 to 0.30 (moderate), and >0.30 (strong) [72]. Significant associations were entered in multivariable linear regression models to examine the association between psychological variables and UTAUT, controlling for sex, age, and smart technology experience. Depressive symptoms (PHQ-9) and generalized anxiety symptoms (GAD-7) were included to evaluate whether PPWB (BIT) predicted variance in SR acceptability independently of psychological distress. Urbanicity and CCI were included as covariates because they were relevant predictors of acceptability in prior analyses of this cohort [31]. Significance was set at *P*=.05. Unstandardized B coefficients with 95% CIs, *t* values, and effect sizes (R^2 and adjusted R^2) were reported, along with variance inflation factor and tolerance statistics to assess multicollinearity. Model fit was evaluated using *F*-statistics and corresponding *P* values for each imputed dataset, as pooled omnibus test statistics are not provided in SPSS for multiply imputed data. Effect sizes were interpreted as small (R^2=0.02), medium (R^2=0.13), and large (R^2=0.26) based on established guidelines [73,74].

In post hoc analyses, we conducted Pearson correlations between each BIT item and UTAUT behavioral intention to identify the item most strongly driving the association. This approach was used because the American Heart Association statement [43] recommends analyzing specific PPWB components, while the BIT provides only a summary score. Additionally, we examined exploratory Pearson correlations between individual BIT items and 25 rated SR capabilities and features to explore potential links between distinct aspects of psychological well-being and feature-specific acceptability, thereby generating preliminary insights relevant to SR design. Given the exploratory nature of these analyses and the large number of comparisons, they were not intended for formal significance testing. Finally, we repeated the multiple regression analyses using HF type (HFrEF and HF with preserved ejection fraction), LVEF, and NYHA class as alternative indicators of HF symptom burden, instead of the CCI, to explore their potential influence on the association between BIT and UTAUT behavioral intention.

XSL•FO
RenderX

Results

Sample Description

The study included 101 patients with HF (women: n=36, 36% female; and White: n=90, 89%), with a mean age of 68 (SD 10) years, and 33% (n=33) resided in urban communities (Table 1 and Multimedia Appendix 2). The average LVEF was 37.0 (SD 13.1), and the majority of patients were diagnosed with HFrEF (77/98, 76%) and classified as NYHA class II (77/101, 76%). The mean comorbidity index was 5.5 (SD 2.0), and the average time since diagnosis was 5.7 (SD 10.0) years.

XSL•FO
RenderX

Table . Participant characteristics of adults with heart failure (N=101).

Variables	Values
Age, mean (SD)	68.0 (9.9)
Female, n (%)	36 (35.6)
Race[g], n (%)	
White	90 (89.1)
Arab	2 (2.0)
Black	2 (2.0)
Aboriginal	1 (1.0)
Chinese	1 (1.0)
Filipino	1 (1.0)
South Asian	1 (1.0)
Other	2 (2.0)
Marital status, n (%)	
Married or partner	70 (69.3)
Divorced or separated	10 (9.9)
Widowed	6 (5.9)
Single	15 (14.9)
Highest education, n (%)	
High school	23 (22.8)
Trade or college degree	37 (36.6)
Bachelor’s	28 (27.7)
Master’s or doctorate	13 (12.9)
Employment status, n (%)	
Employed	13 (12.9)
Unemployed, disability, or other	18 (17.9)
Retired	70 (69.3)
Community, n (%)	
Rural	28 (27.7)
Urban	33 (32.7)
Suburban	33 (32.7)
Mixed	7 (6.9)
Would likely use SR[b], n (%)	67 (69.1)
Smart devices in use, mean (SD)	4.5 (2.2)
HFrEF[c], n (%)	77 (76.2)
LVEF[d], mean (SD)	37.0 (13.1)
NYHA[e] class II, n (%)	77 (76.2)
Years since diagnosis, mean (SD)	5.7 (10.0)
CCI[f], mean (SD)	5.5 (2.1)

[a]Race data were missing for 1 participant.

[b]SR: social robot.

[c]HFrEF: heart failure with reduced ejection fraction.

[d]LVEF: left ventricular ejection fraction.

XSL•FO
RenderX

[e]NYHA: New York Heart Association.

[f]CCI: Charlson Comorbidity Index.

Experience With Technology

Most of the 101 participants reported using at least one smart device, including smartphones (n=87, 86%), laptops (n=69, 68%), iPads or tablets (n=54, 54%), or smart televisions (n=57, 56%). Fewer patients used wearable devices, such as smartwatches (n=31, 31%), voice assistants (eg, Google Home and Alexa; n=30, 30%), and smart medical or health-related sensors (eg, respiratory rate monitors, sleep monitors, heart rate sensors, and fall detection; n=21, 21%). The smart technology experience, indicated by the mean number of smart devices reported, was 4.5 (SD 2.2; range 0 - 12).

General Attitudes Toward SRs

As previously reported, more than half of the participants (53/99, 54%) reported being not at all or only somewhat aware of SRs before participating in this study, and 93% (92/99) had never interacted with an SR in the past [31]. In terms of usefulness, 58% (57/98) agreed or somewhat agreed that an SR would be helpful in their current health situation, and 69% (67/97) agreed or somewhat agreed that they would use an SR if one were available to them. Additional information regarding the samples' preferred SR capabilities may be found elsewhere [31].

Psychological, Personality, and Technology Acceptance Measures

The mean PHQ-9 score was 6.4 (SD 5.5), with 23% (23/101) of patients scoring in the clinical range (ie, above the cutoff for moderate depressive symptoms). The mean GAD-7 score was 4.8 (SD 5.1), with 17% (17/101) of patients crossing the threshold for clinically significant anxiety symptoms. Patients reported a moderate BIT score (mean 36.9, SD 7.1). Among the personality traits, conscientiousness (mean 11.1, SD 2.4) and agreeableness (mean 10.7, SD 2.3) received the highest ratings, followed by emotional stability (mean 10.1, SD 3.0), openness to experiences (mean 9.8, SD 2.5), and extraversion (mean 7.8, SD 2.8). Scores on the UTAUT subscales showed moderate acceptability of SRs in terms of performance expectancy (mean 17.0, SD 7.3; scale range 4 - 28), social influence (mean 15.4, SD 5.6; scale range 4 - 28), and facilitating conditions (mean 20.6, SD 4.2; scale range 4 - 28). Acceptability was high regarding effort expectancy (mean 21.4, SD 5.6; scale range 4 - 28). The behavioral intention to use SRs was moderate (mean 13.7, SD 4.9; scale range 3 - 21; Table 2).

XSL•FO
RenderX

Table . Correlation matrix (Pearson *r* and two tailed *P*-value) of questionnaire scores assessing technology acceptability, mental health symptoms, psychological well-being, and personality traits in adults with heart failure (N=101).[a]

Variable	UTAUT[b] performance	UTAUT effort	UTAUT conditions	UTAUT social	UTAUT behavioral intention	PHQ-9[c]	GAD-7[d]	BIT[e]	TIPI[f] – Extraversion	TIPI – Agreeableness	TIPI – Conscientiousness	TIPI - Emotional Stability	TIPI - Openness to Experiences
UTAUT performance													
r	1	.505	.383	.684	.783	.164	.170	.083	-.072	-.071	-.050	-.104	-.035
P value	—	<.001	<.001	<.001	<.001	.11	.09	.41	.48	.49	.63	.30	.74
UTAUT effort													
r	.505	1	.745	.448	.597	-.090	-.060	.234	-.077	.124	-.013	.033	.069
P value	<.001	—	<.001	<.001	<.001	.38	.57	.02	.45	.25	.90	.74	.55
UTAUT conditions													
r	.383	.745	1	.337	.518	-.045	-.057	.263	.019	.033	-.024	-.043	.144
P value	<.001	<.001	—	.002	<.001	.66	.58	.008	.85	.76	.83	.69	.22
UTAUT social													
r	.684	.448	.337	1	.634	.013	.047	.238	-.029	.012	-.023	.080	-.049
P value	<.001	<.001	.002	—	<.001	.90	.65	.03	.78	.91	.82	.45	.65
UTAUT behavioral intention													
r	.783	.597	.518	.634	1	.031	.136	.198	-.026	-.015	-.001	-.057	.003
P value	<.001	<.001	<.001	<.001	—	.77	.21	.049	.80	.89	.99	.59	.98
PHQ-9													
r	.164	-.090	-.045	.013	.031	1	.756	-.554	-.257	-.238	-.388	-.551	-.136
P value	.11	.38	.66	.90	.77	—	<.001	<.001	.01	.02	<.001	<.001	.18
GAD-7													
r	.170	-.060	-.057	.047	.136	.756	1	-.399	-.237	-.196	-.237	-.629	-.193
P value	.09	.57	.58	.65	.21	<.001	—	<.001	.02	.05	.02	<.001	.06
BIT													
r	.083	.234	.263	.238	.198	-.554	-.399	1	.231	.264	.330	.349	.199
P value	.41	.02	.008	.03	.049	<.001	<.001	—	.02	.01	.001	<.001	.048
TIPI – Extraversion													
r	-.072	-.077	.019	-.029	-.026	-.257	-.237	.231	1	.032	.153	.212	.182

XSL•FO
RenderX

Variable	UTAUT[b] performance	UTAUT effort	UTAUT conditions	UTAUT social	UTAUT behavioral intention	PHQ-9[c]	GAD-7[d]	BIT[e]	TIPI[f] – Extraversion	TIPI – Agreeableness	TIPI – Conscientiousness	TIPI - Emotional Stability	TIPI - Openness to Experiences
P value	.48	.45	.85	.78	.80	.01	.02	.02	—	.76	.13	.04	.07
TIPI – Agreeableness													
r	-.071	.124	.033	.012	-.015	-.238	-.196	.264	.032	1	.230	.490	.259
P value	.49	.25	.76	.92	.89	.02	.05	.01	.76	—	.02	<.001	.01
TIPI – Conscientiousness													
r	-.050	-.013	-.024	-.023	-.001	-.388	-.237	.330	.153	.230	1	.314	.248
P value	.63	.90	.83	.82	.99	<.001	.02	.001	.13	.02	—	.002	.02
TIPI - Emotional Stability													
r	-.104	.033	-.043	.080	-.057	-.551	-.629	.349	.212	.490	.314	1	.242
P value	.30	.74	.69	.45	.59	<.001	<.001	<.001	.04	<.001	.002	—	.02
TIPI - Openness to Experiences													
r	-.035	.069	.144	-.049	.003	-.136	-.193	.199	.182	.259	.248	.242	1
P value	.74	.55	.22	.65	.98	.18	.06	.048	.07	.01	.02	.02	—

[a]Data were multiply imputed and pooled. Higher scores indicate greater levels of the respective construct.

[b]UTAUT: Unified Theory of Acceptance and Use of Technology.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]GAD-7: Generalized Anxiety Disorder-7.

[e]BIT: Brief Inventory of Thriving.

[f]TIPI: Ten-Item Personality Inventory.

Association of Psychological Variables With SR Acceptability

Table 2 displays the means and correlations between the study variables. PPWB (BIT score) was positively correlated with UTAUT effort expectancy ($r(99)$=0.234; P=.02), UTAUT facilitating conditions ($r(99)$=0.263; P=.008), UTAUT social influence ($r(99)$=0.238; P=.03), and UTAUT behavioral intention ($r(99)$=0.198; P=.049). There were no significant correlations between PHQ-9, GAD-7, and TIPI with any of the UTAUT subscales.

Across the original and imputed datasets, the overall regression model predicting performance expectancy was not statistically significant ($F_{8,77\text{-}90}$=1.24-1.55; all P>.15). The model predicting social influence was statistically significant in the original dataset ($F_{8,75}$=2.21; P=.04) but not in any of the imputed datasets ($F_{8,90}$=1.12-1.91; all P≥.07). In contrast, the models predicting effort expectancy, facilitating conditions, and behavioral intention were statistically significant across the original and all imputed datasets (effort expectancy: $F_{8,77\text{-}90}$=3.11 - 4.29; all P≤.004; facilitating conditions: $F_{8,76\text{-}90}$=3.13 - 3.66; all P≤.004; and behavioral intention: $F_{8,79\text{-}90}$=2.20 - 3.12; all P≤.04).

In the multiple regression models (Tables 3 and 4, Figure 1, Supplement S3-S6 in Multimedia Appendix 1), higher BIT scores were associated with increased SR acceptability, including UTAUT facilitating conditions (B=0.17; P=.01) and behavioral intention (B=0.17; P=.04). All models were adjusted

XSL•FO
RenderX

for sex, age, smart technology experience, urbanicity, CCI, depressive, and anxiety symptoms.

Table . Multiple regression analyses examining the association of psychological well-being (Brief Inventory of Thriving [BIT]) with technology acceptability (Unified Theory of Acceptance and Use of Technology [UTAUT]) adjusted for covariates in adults with heart failure (N=101).[a]

	UTAUT performance expectancy						UTAUT effort expectancy						UTAUT facilitating conditions					
Predictors	B (SE)	95% CI	*t* [b]	*P* value	Tolerance	VIF[c]	B (SE)	95% CI	*t*	*P* value	Tolerance	VIF	B (SE)	95% CI	*t*	*P* value	Tolerance	VIF
Sex[d]	-0.64 (1.60)	-3.78, 2.50	-0.4	.9	0.91	1.1	-0.59 (1.09)	-2.72, 1.54	-0.54	.59	0.93	1.07	-0.63 (0.82)	-2.25, 0.98	-0.77	.44	0.92	1.08
Age	0.00 (0.09)	-0.17, 0.17	-0.03	.98	0.72	1.4	-0.14 (0.06)	-0.26, -0.02	-2.34	.02	0.71	1.41	-0.05 (0.05)	-0.14, 0.04	-1	.32	0.73	1.37
Smart technology experience[e]	0.36 (0.35)	-0.32, 1.04	1.03	.30	0.86	1.16	0.45 (0.24)	-0.03, 0.93	1.85	.06	0.87	1.15	0.48 (0.18)	0.12, 0.84	2.63	.009	0.87	1.16
Urbanicity[f]	1.13 (1.63)	-2.07, 4.32	0.69	.49	0.94	1.07	-1.23 (1.11)	-3.40, 0.94	-1.11	.26	0.93	1.07	-0.18 (0.88)	-1.91, 1.55	-0.2	.84	0.93	1.08
CCI[g]	-0.66 (0.43)	-1.49, 0.18	-1.55	.12	0.69	1.44	-0.24 (0.29)	-0.81, 0.33	-0.82	.41	0.69	1.46	-0.41 (0.22)	-0.85, 0.02	-1.88	.06	0.71	1.41
PHQ-9[h]	0.27 (0.24)	-0.20, 0.73	1.12	.26	0.34	2.91	-0.11 (0.16)	-0.43, 0.20	-0.7	.49	0.35	2.86	0.08 (0.12)	-0.16, 0.32	0.65	.52	0.34	2.94
GAD-7[i]	0.12 (0.22)	-0.31, 0.55	0.54	.59	0.42	2.36	0.02 (0.16)	-0.29, 0.32	0.11	.92	0.44	2.28	-0.05 (0.12)	-0.29, 0.18	-0.43	.66	0.42	2.37
BIT	0.23 (0.13)	-0.02, 0.48	1.84	0.07	0.63	1.58	0.16 (0.09)	-0.02, 0.33	1.75	.08	0.64	1.57	0.17 (0.07)	0.04, 0.30	2.49	.01	0.64	1.56

[a]Results are pooled estimates from multiple imputation; therefore, unstandardized regression coefficients (B) are reported instead of standardized betas (β). Each B represents the expected change in the outcome score for a 1-unit increase in the predictor. Model fit (F-statistics and P values) varied across imputed datasets and is therefore summarized in the text rather than pooled in the table, as SPSS does not provide pooled omnibus test statistics.

[b]Degrees of freedom for t values are not reported in the table, as they are estimated under Rubin's rules for multiple imputation and vary across imputations.

[c] VIF: variance inflation factor.

[d]Sex: 0=male, 1=female.

[e]Smart technology experience=number of devices in use.

[f]Urbanicity: 0=nonurban and 1=urban.

[g]CCI: Charlson Comorbidity Index.

[h] PHQ-9: Patient Health Questionnaire-9.

[i]GAD-7: Generalized Anxiety Disorder-7.

XSL•FO
RenderX

Table . Multiple regression analyses examining the association of psychological well-being (Brief Inventory of Thriving [BIT]) with technology acceptability (Unified Theory of Acceptance and Use of Technology [UTAUT]) adjusted for covariates in adults with heart failure (N=101).[a]

	UTAUT social influence						UTAUT behavioral intention					
Predictors	B (SE)	95% CI	*t* [b]	*P* value	Tolerance	VIF[c]	B (SE)	95% CI	*t*	*P* value	Tolerance	VIF
Sex[d]	-0.33 (1.19)	-2.67, 2.01	-0.28	.78	0.92	1.08	-0.04 (0.99)	-1.98, 1.91	-0.04	.97	0.92	1.09
Age	0.10 (0.07)	-0.03, 0.23	1.49	.14	0.72	1.38	0.01 (0.06)	-0.10, 0.12	0.18	.86	0.73	1.38
Smart technology experience[e]	-0.04 (0.27)	-0.57, 0.48	-0.17	.87	0.86	1.17	0.53 (0.23)	0.08, 0.99	2.3	.02	0.87	1.15
Urbanicity[f]	-0.07 (1.20)	-2.42, 2.28	-0.06	.95	0.93	1.08	0.07 (1.04)	-1.97, 2.12	0.07	.94	0.93	1.08
CCI[g]	-0.29 (0.32)	-0.91, 0.34	-0.89	.37	0.7	1.42	-0.50 (0.27)	-1.03, 0.03	-1.84	.07	0.7	1.42
PHQ-9[h]	0.17 (0.18)	-0.18, 0.52	0.95	.34	0.33	3	-0.06 (0.15)	-0.36, 0.24	-0.37	.71	0.34	2.92
GAD-7[i]	0.06 (0.17)	-0.26, 0.39	0.38	.70	0.41	2.47	0.23 (0.15)	-0.06, 0.52	1.57	.12	0.43	2.35
BIT	0.28 (0.10)	0.07, 0.48	2.8	.008	0.65	1.54	0.17 (0.09)	0.00, 0.34	2.02	.04	0.64	1.57

[a]Results are pooled estimates from multiple imputation; therefore, unstandardized regression coefficients (B) are reported instead of standardized betas (β). Each B represents the expected change in the outcome score for a 1-unit increase in the predictor. Model fit (F-statistics and P values) varied across imputed datasets and is therefore summarized in the text rather than pooled in the table, as SPSS does not provide pooled omnibus test statistics.

[b]Degrees of freedom for *t* values are not reported in the table, as they are estimated under Rubin's rules for multiple imputation and vary across imputations.

[c]VIF: variance inflation factor.

[d]Sex: 0=male, 1=female.

[e]Smart technology experience=number of devices in use.

[f]Urbanicity: 0=nonurban and 1=urban.

[g]CCI: Charlson Comorbidity Index.

[h]PHQ-9: Patient Health Questionnaire-9.

[i]GAD-7: Generalized Anxiety Disorder-7.

XSL•FO
RenderX

Figure 1. Forest plot of regression coefficients (B) and 95% CI for predictors of Unified Theory of Acceptance and Use of Technology behavioral intention. Sex: 0=male, 1=female; Smart technology experience=number of devices in use; Urbanicity: 0=nonurban, 1=urban; BIT: Brief Inventory of Thriving; CCI: Charlson Comorbidity Index; GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9.



In addition, 2 covariates showed significant associations with UTAUT subscales. Age was negatively associated with effort expectancy (B=−0.14; P=.02). Smart technology experience was positively associated with facilitating conditions (B=0.48; P=.009) and behavioral intention (B=0.53; P=.02).

Post Hoc Analyses

Among the 10 items of the BIT, UTAUT behavioral intention to use an SR was positively correlated with "I am achieving most of my goals" (r(99)=0.233; P=.02) and "I am optimistic about my future" (r(99)=0.210; P=.04), while correlations with "My life has a clear sense of purpose" (r(99)=.189; P=.07), "What I do in life is valuable and worthwhile" (r(99)=0.184; P=.07), "I can succeed if I put my mind to it" (r(99)=0.142; P=.17), "I feel good most of the time" (r(99)=0.139; P=.17), "My life is going well" (r(99)=0.108; P=.29), "In most activities I do, I feel energized" (r(99)=0.092; P=.37), "I feel a sense of belonging in my community" (r(99)=0.124; P=.24), and "There are people who appreciate me as a person" (r(99)=0.073; P=.48) were small and not statistically significant.

The BIT total score was associated with the SR feature "Entertainment" (r(93)=0.216). Among individual items, "I feel a sense of belonging to my community" was linked to the largest number of SR features (Supplement S7 in Multimedia Appendix 1). Including HF type, LVEF, or NYHA class as control variables for symptom burden (instead of CCI) in the multiple regression analyses did not affect the association between BIT and UTAUT behavioral intention.

Discussion

Principal Findings

This study is a pioneering study to investigate psychological factors associated with determinants of SR acceptability in an HF patient sample. The findings highlight that PPWB is associated with dimensions of patients' SR acceptability, independent of psychological distress: patients with higher PPWB are more likely to feel assured in their personal resources and knowledge to operate SRs and more likely to intend to use an SR in the future. These findings extend our earlier research, which demonstrated that urbanicity and lower comorbidity levels are linked to higher SR acceptability [31], positioning PPWB as a key psychological factor, alongside demographic and clinical variables, in influencing patients' willingness to adopt SRs. This is consistent with broader HF literature linking PPWB to greater adherence to health behaviors [41,42] and increased self-care [75], underscoring the relevance of psychological well-being in HF management. Taken together, these findings suggest that patients with higher PPWB may not only be more motivated to manage their condition actively but also more receptive to emerging tools that support autonomy and self-efficacy in care.

XSL•FO
RenderX

In particular, the BIT items, "I am achieving most of my goals" and "I am optimistic about my future" exhibited the strongest associations with patients' intention to use an SR. This indicates that a forward-looking, goal-oriented mindset may increase openness to innovative health care tools such as SRs. In line with this, a previous study found that satisfaction with life was associated with higher SR acceptability in healthy adults [34]. Moreover, research in general populations has indicated that positive emotions, such as enjoyment and excitement, are linked to a higher willingness to engage with SRs [36,76] and a higher confidence regarding one's ability to use SRs [38]. In our HF population, an optimistic outlook on life seems to be a more significant predictor of SR acceptability than momentary emotional states. This may be explained by our participants not directly interacting with SRs in a trial setting but instead simply expressing their attitudes toward SRs. Further research is needed to explore how different PPWB components influence SR acceptability in HF populations.

Interestingly, although 23% (23/101) of participants had at least moderate depressive symptoms and 17% (17/101) at least clinically elevated anxiety symptoms, neither depression nor anxiety scores were significantly associated with determinants of SR acceptability. This finding is consistent with a study in healthy older adults, where depressive symptoms were not associated with SR acceptability [34]. Similarly, in students, negative emotions such as fear and anxiety were either weakly or not at all related to SR acceptability or willingness to interact with SRs in laboratory settings [36]. While some research has linked anxiety to lower perceived ease of use of SRs [38], we found no such relationship. Conversely, studies examining loneliness, which is often associated with depressive symptoms [77], have found it to be positively correlated with SR acceptability, particularly during the COVID-19 pandemic [78]. Notably, in our prior study, few patients with HF expressed interest in companionship features; instead, most were drawn to SRs for functional tasks such as monitoring (eg, blood pressure tracking) and managing appointments [31]. This suggests that SR features focusing on reminders, monitoring, and educational support may be more appealing to patients with HF than those designed for social interaction or mental health.

Unexpectedly, personality traits, such as agreeableness, extraversion, and openness, did not show a significant association with SR acceptability in our HF sample. In contrast, Esterwood et al [51] reported in their meta-analysis that individuals who are more agreeable, extroverted, and open are more likely to accept SRs, but they also acknowledged that this link weakens in older populations. Our sample included older adults, and if age indeed moderates these associations (with weaker associations between personality and SR acceptability with age) as suggested by the findings of Esterwood et al [51] and Kim et al [79], this may help understand our null effect. Alternatively, the low internal consistency of the 2-item TIPI subscales in our sample (α=.33-.40) may have contributed to measurement error, and the null associations with personality traits should be interpreted with caution.

Notably, the general acceptability of SRs in this sample was high, with 69% (67/97) of participants indicating they would likely use an SR if available. This level is similar to a small sample of patients in cardiac rehabilitation [29]. This suggests that SR technology is acceptable in patients with cardiac conditions. Additionally, our investigation into HF health care providers' perspectives similarly revealed openness to SRs in roles such as data collection, education, and patient prompts but highlighted hesitancy regarding interventional applications [30].

In light of these findings, when considering the integration of SRs into HF care, clinicians should be aware that patients' PPWB, particularly their sense of optimism and goal-directedness, may significantly influence acceptance. Assessing PPWB may help identify patients most receptive to such technologies and guide tailored implementation strategies. Co-designing SR features in collaboration with patients, health care providers, and developers, guided by intervention mapping [80,81], can ensure that SRs are functionally relevant and emotionally engaging. Our previous findings suggest that patients with HF living in urban areas with low comorbidities may be ideal candidates for testing an initial SR prototype [31]. This study further indicates that patients with high PPWB may be more likely to engage with and benefit from SRs. In particular, SR features such as self-tracking, proactive health alerts, and especially gamification could empower these individuals to maintain their well-being [82]. Consistent with this, the PPWB score was associated with the SR feature "Entertainment," highlighting the potential motivational role of enjoyable, engaging interactions. The fact that "I feel a sense of belonging in my community" was linked to the largest number of SR features indicates that social connectedness may be a key driver of engagement across multiple SR functions.

Our findings also suggest that individual characteristics related to age and technology familiarity may influence patient acceptability of SRs. Specifically, higher perceived effort among older participants highlights potential age-related barriers to adoption. In contrast, greater smart technology experience appears to facilitate both perceived resources and intentions to use. These findings suggest that SRs should be designed to minimize complexity and cognitive load to better support older users, while also incorporating intuitive interfaces, clear guidance, and accessible support features to accommodate individuals with limited prior experience with digital technologies. For patients experiencing emotional distress or lower levels of positive affect, SRs may benefit from simplified, nonintrusive designs that provide instrumental support without adding cognitive burden [9]. For patients who are skeptical of SRs, controlled environments where they can test the technology could help alleviate concerns and build trust [83]. Research has shown that hands-on interaction with SRs can significantly increase acceptance of the technology [84]. On the basis of our findings, relevant features for patients appear to include monitoring and appointment scheduling, which could be prioritized for initial testing [31].

Future research should explore the longitudinal effects of SR use on emotional well-being and health outcomes, providing crucial insights into the sustainability and broader impact of SR integration in HF care. Longitudinal studies will also be essential in tracking how SR acceptance evolves over time and throughout the course of illness, identifying the key determinants of

XSL•FO
RenderX

sustained use, which is a central challenge in SR applications [85]. By advancing collaborative and patient-centered design processes, SR technologies can be optimized not only to meet the functional demands of HF management but also to foster meaningful patient engagement.

Limitations

First, the relatively small sample size may limit the generalizability of our findings to the broader HF population. In our sample, only 36% (36/100) were women, which is consistent with previous HF studies [86], but slightly lower than the expected 41% (41/100) of women in the overall HF population. While prior research suggests that men are generally more likely than women to engage with technology and report higher technology acceptance in older adult populations [87], these differences appear to be minimal or may even be reversed in the context of SRs [88,89]. As such, the underrepresentation of women may influence the generalizability of the acceptability estimates observed in this study, although the direction and magnitude of any such effect remain uncertain. The sample was also predominantly White (90/101, 90%), which may limit generalizability across ethnic and cultural groups [90]. Structural, social, and cultural factors, such as access to technology, familiarity with digital tools, and culturally shaped attitudes toward robots, could influence both the perceived ease of use and the support available for technology adoption, potentially altering the relationships observed in this study. These considerations underscore the need for more inclusive recruitment strategies to ensure that the perspectives and needs of diverse populations are adequately represented in future research [91].

Second, the predominance of NYHA class II patients in our sample raises the possibility that individuals with more advanced HF and greater symptom burden may experience lower levels of PPWB, such as reduced optimism or sense of control, which could in turn affect their openness to adopting SR technology. Patients with more severe HF may also face greater challenges in using technology, perceive higher effort to engage with SRs, and require additional support for adoption. However, *post hoc* analyses revealed that HF type, LVEF, and NYHA class did not significantly influence the relationship between PPWB and SR acceptability, suggesting that the observed associations may be consistent across different clinical presentations.

Third, the time since diagnosis showed a right-skewed distribution, which should be considered when interpreting summary statistics. Furthermore, the TIPI subscales showed low internal consistency (α=.33-.40), which may have contributed to measurement error and warrants cautious interpretation of null findings. While the TIPI was chosen for brevity and minimal participant burden, future studies could use more comprehensive personality assessments.

Fourth, the videos patients watched before completing the questionnaires may have influenced their responses, potentially creating a framing effect, where marketing-style presentation could inflate perceptions of performance expectancy and minimize perceived effort expectancy compared to real-world use. However, this introduction was essential, as 93% (92/99) of patients were unfamiliar with SRs. As this was early-phase research with limited funding, we opted not to provide a demonstration of a physical robot. It is possible that patients' perspectives on using SRs might have been different if they had the opportunity to interact directly with an SR.

Finally, as the study used a cross-sectional design, temporal relationships cannot be discerned. While emotional states such as PPWB may be key predictors of SR acceptability, it remains unclear how these factors influence long-term usage. Longitudinal studies could assess whether patients with high PPWB not only accept SRs more readily but also maintain consistent engagement over time. Importantly, this study is the first to investigate the association between psychological factors and SR acceptability among patients with HF, setting the stage for future longitudinal research and the co-design of SR interventions tailored to the unique needs of HF care.

Conclusions

This study highlights the role of PPWB in relation to determinants of SR acceptability among patients with HF, while negative psychological factors and personality traits showed no significant associations. These findings emphasize the importance of considering PPWB when designing and implementing SR interventions in HF care. Patients with high PPWB may be ideal candidates for testing an initial SR prototype, as they are more likely to engage with and benefit from SRs.

Acknowledgments

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) and Microsoft Copilot to assist with language editing and clarity. All artificial intelligence–generated suggestions were reviewed and revised by the authors, who take full responsibility for the final content.

Funding

This work was supported by a grant from the Ottawa Region for Advanced Cardiovascular Research Excellence (ORACLE), a Brain-Heart Interconnectome Canada First Research Excellence Fund and a Brain-Heart Research Integrative Innovation Team Endeavor Team Grant from Heart & Stroke Foundation.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

LMM was responsible for formal analysis, investigation, visualization, writing the original draft, and reviewing and editing. KB, JGF, KD, MG, TC, CM, and PPL contributed to conceptualization, funding acquisition, investigation, methodology, and reviewing and editing. HT was responsible for conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, writing the original draft, and reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video descriptions and transcripts, study questionnaires, and additional tables and figures supporting the main analyses.
[PDF File, 1081 KB - cardio_v10i1e83163_app1.pdf]

Multimedia Appendix 2
Graphical abstract.
[PNG File, 290 KB - cardio_v10i1e83163_app2.png]

References

1. Khan MS, Shahid I, Bennis A, Rakisheva A, Metra M, Butler J. Global epidemiology of heart failure. Nat Rev Cardiol 2024 Oct;21(10):717-734. [doi: 10.1038/s41569-024-01046-6] [Medline: 38926611]
2. Desai AS, Stevenson LW. Rehospitalization for heart failure. Circulation 2012 Jul 24;126(4):501-506. [doi: 10.1161/CIRCULATIONAHA.112.125435]
3. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
4. Moradi M, Daneshi F, Behzadmehr R, Rafiemanesh H, Bouya S, Raeisi M. Quality of life of chronic heart failure patients: a systematic review and meta-analysis. Heart Fail Rev 2020 Nov;25(6):993-1006. [doi: 10.1007/s10741-019-09890-2] [Medline: 31745839]
5. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. Circulation 2019 Mar 5;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
6. Bozkurt B, Ahmad T, Alexander K, et al. HF STATS 2024: heart failure epidemiology and outcomes statistics an updated 2024 report from the Heart Failure Society of America. J Card Fail 2025 Jan;31(1):66-116. [doi: 10.1016/j.cardfail.2024.07.001] [Medline: 39322534]
7. Sheridan TB. A review of recent research in social robotics. Curr Opin Psychol 2020 Dec;36:7-12. [doi: 10.1016/j.copsyc.2020.01.003] [Medline: 32294577]
8. Dautenhahn K. Methodology & themes of human-robot interaction: a growing research field. International Journal of Advanced Robotic Systems 2007 Mar 1;4(1):15. [doi: 10.5772/5702]
9. Breazeal C, Dautenhahn K, Kanda T. Social robotics. In: Siciliano B, Khatib O, editors. Springer Handbook of Robotics: Springer; 2016:1935-1972. [doi: 10.1007/978-3-319-32552-1_72]
10. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open 2018 Feb 12;8(2):e018815. [doi: 10.1136/bmjopen-2017-018815] [Medline: 29440212]
11. Robinson H, MacDonald B, Broadbent E. The role of healthcare robots for older people at home: a review. Int J of Soc Robotics 2014 Nov;6(4):575-591. [doi: 10.1007/s12369-014-0242-2]
12. Diamond J, DeVore AD. New strategies to prevent rehospitalizations for heart failure. Curr Treat Options Cardiovasc Med 2022;24(12):199-212. [doi: 10.1007/s11936-022-00969-y] [Medline: 36164396]
13. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. Gerontologist 2019 Jan 9;59(1):e37-e51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
14. Broadbent E, Garrett J, Jepsen N, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45. [doi: 10.2196/jmir.8640] [Medline: 29439942]
15. Chen SC, Jones C, Moyle W. Social robots for depression in older adults: a systematic review. J Nurs Scholarsh 2018 Nov;50(6):612-622. [doi: 10.1111/jnu.12423] [Medline: 30198077]
16. Kanamori M, Suzuki M, Oshiro H, et al. Pilot study on improvement of quality of life among elderly using a pet-type robot. Presented at: 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation; Jul 16-20, 2003; Kobe, Japan p. 107-112. [doi: 10.1109/CIRA.2003.1222072]
17. Mervin MC, Moyle W, Jones C, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. J Am Med Dir Assoc 2018 Jul;19(7):619-622. [doi: 10.1016/j.jamda.2017.10.008] [Medline: 29325922]

XSL•FO
RenderX

18. Robinson NL, Cottier TV, Kavanagh DJ. Psychosocial health interventions by social robots: systematic review of randomized controlled trials. J Med Internet Res 2019 May 10;21(5):e13203. [doi: 10.2196/13203] [Medline: 31094357]
19. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
20. Budak KB, Atefi G, Hoel V, et al. Can technology impact loneliness in dementia? A scoping review on the role of assistive technologies in delivering psychosocial interventions in long-term care. Disabil Rehabil Assist Technol 2023 Oct;18(7):1107-1119. [doi: 10.1080/17483107.2021.1984594] [Medline: 34752177]
21. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 1;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
22. Moradi M, Doostkami M, Behnamfar N, Rafiemanesh H, Behzadmehr R. Global prevalence of depression among heart failure patients: a systematic review and meta-analysis. Curr Probl Cardiol 2022 Jun;47(6):100848. [doi: 10.1016/j.cpcardiol.2021.100848] [Medline: 34006389]
23. Pino M, Boulay M, Jouen F, Rigaud AS. "Are we ready for robots that care for us?" attitudes and opinions of older adults toward socially assistive robots. Front Aging Neurosci 2015;7:141. [doi: 10.3389/fnagi.2015.00141] [Medline: 26257646]
24. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J of Soc Robotics 2018 Nov;10(5):643-668. [doi: 10.1007/s12369-018-0471-x]
25. de Graaf MMA, Ben Allouch S, van Dijk J. Why would I use this in my home? A model of domestic social robot acceptance. Hum Comput Interact 2019 Mar 4;34(2):115-173. [doi: 10.1080/07370024.2017.1312406]
26. Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: the Almere Model. Int J of Soc Robotics 2010 Dec;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
27. Naneva S, Sarda Gou M, Webb TL, Prescott TJ. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. Int J of Soc Robotics 2020 Dec;12(6):1179-1201. [doi: 10.1007/s12369-020-00659-4]
28. Casas J, Senft E, Gutiérrez LF, et al. Social assistive robots: assessing the impact of a training assistant robot in cardiac rehabilitation. Int J of Soc Robotics 2021 Sep;13(6):1189-1203. [doi: 10.1007/s12369-020-00708-y]
29. Casas JA, Céspedes N, Cifuentes CA, Gutierrez LF, Rincón-Roncancio M, Múnera M. Expectation vs. reality: attitudes towards a socially assistive robot in cardiac rehabilitation. Appl Sci (Basel) 2019;9(21):4651. [doi: 10.3390/app9214651]
30. Bouchard K, Liu PP, Dautenhahn K, et al. Cardiology professionals' views of social robots in augmenting heart failure patient care. Eur Heart J Digit Health 2024 Jan;5(1):69-76. [doi: 10.1093/ehjdh/ztad067] [Medline: 38264699]
31. Maukel LM, Knight L, Bouchard K, et al. The role of social robots in heart failure: exploring patient attitudes, acceptability, and needs. JACC Heart Fail 2025 Sep;13(9):102570. [doi: 10.1016/j.jchf.2025.102570] [Medline: 40680705]
32. Feltman R, Elliot AJ. Approach and avoidance motivation. In: Seel NM, editor. Encyclopedia of the Sciences of Learning: Springer; 2012:286-288. [doi: 10.1007/978-1-4419-1428-6_1749]
33. Bishop L, van Maris A, Dogramadzi S, Zook N. Social robots: the influence of human and robot characteristics on acceptance. Paladyn, Journal of Behavioral Robotics 2019 Jan 1;10(1):346-358. [doi: 10.1515/pjbr-2019-0028]
34. Baisch S, Kolling T, Schall A, et al. Acceptance of social robots by elder people: does psychosocial functioning matter? Int J of Soc Robotics 2017 Apr;9(2):293-307. [doi: 10.1007/s12369-016-0392-5]
35. Luo C, Yuan R, Mao B, Liu Q, Wang W, He Y. Technology acceptance of socially assistive robots among older adults and the factors influencing it: a meta-analysis. J Appl Gerontol 2024 Feb;43(2):115-128. [doi: 10.1177/07334648231202669] [Medline: 37750859]
36. Smith ER, Sherrin S, Fraune MR, Šabanović S. Positive emotions, more than anxiety or other negative emotions, predict willingness to interact with robots. Pers Soc Psychol Bull 2020 Aug;46(8):1270-1283. [doi: 10.1177/0146167219900439] [Medline: 31959093]
37. Fredrickson BL. The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. Am Psychol 2001 Mar;56(3):218-226. [doi: 10.1037//0003-066x.56.3.218] [Medline: 11315248]
38. de Graaf MMA, Ben Allouch S. Exploring influencing variables for the acceptance of social robots. Rob Auton Syst 2013 Dec;61(12):1476-1486. [doi: 10.1016/j.robot.2013.07.007]
39. Hoong ALS, Thi LS, Lin MH. Affective technology acceptance model: extending technology acceptance model with positive and negative affect. In: Mohiuddin M, Halilem N, Ahasanul Kobir SM, Yuliang C, editors. Knowledge Management Strategies and Applications: IntechOpen; 2017:147-165. [doi: 10.5772/intechopen.70351]
40. Maukel LM, Bouchard K, Liu PP, et al. The heart's silver lining: positive psychological well-being and health related quality of life in patients with heart failure. Gen Hosp Psychiatry 2025;95:11-18. [doi: 10.1016/j.genhosppsych.2025.03.014] [Medline: 40187001]
41. Celano CM, Beale EE, Freedman ME, et al. Positive psychological constructs and health behavior adherence in heart failure: a qualitative research study. Nurs Health Sci 2020 Sep;22(3):620-628. [doi: 10.1111/nhs.12704] [Medline: 32125066]
42. Granata N, Torlaschi V, Zanatta F, et al. Positive affect as a predictor of non-pharmacological adherence in older chronic heart failure (CHF) patients undergoing cardiac rehabilitation. Psychol Health Med 2023 Mar;28(3):606-620. [doi: 10.1080/13548506.2022.2077394] [Medline: 35603663]

XSL•FO
RenderX

43. Levine GN, Cohen BE, Commodore-Mensah Y, et al. Psychological health, well-being, and the mind-heart-body connection: a scientific statement from the American Heart Association. Circulation 2021 Mar 9;143(10):e763-e783. [doi: 10.1161/CIR.0000000000000947] [Medline: 33486973]
44. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28. [doi: 10.1159/000353263] [Medline: 24281296]
45. Schimmack U. Methodological issues in the assessment of the affective component of subjective well-being. In: Oxford Handbook of Methods in Positive Psychology: Oxford University Press; 2007:96-110. [doi: 10.1093/oso/9780195172188.003.0007]
46. Widiger TA, Crego C. The five factor model of personality structure: an update. World Psychiatry 2019 Oct;18(3):271-272. [doi: 10.1002/wps.20658] [Medline: 31496109]
47. McCrae RR, Costa Jr PT. The five-factor theory of personality. In: John OP, Robins RW, Pervin LA, editors. Handbook of Personality: Theory and Research: The Guilford Press; 2008:159-181.
48. Elliot AJ, Thrash TM. Approach-avoidance motivation in personality: approach and avoidance temperaments and goals. J Pers Soc Psychol 2002 May;82(5):804-818. [doi: 10.1037//0022-3514.82.5.804] [Medline: 12003479]
49. Jokela M, Pulkki-Råback L, Elovainio M, Kivimäki M. Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies. J Behav Med 2014 Oct;37(5):881-889. [doi: 10.1007/s10865-013-9548-z]
50. Axelsson M, Brink E, Lundgren J, Lötvall J. The influence of personality traits on reported adherence to medication in individuals with chronic disease: an epidemiological study in West Sweden. PLoS One 2011 Mar 28;6(3):e18241. [doi: 10.1371/journal.pone.0018241] [Medline: 21464898]
51. Esterwood C, Essenmacher K, Yang H, Zeng F, Robert LP. A meta-analysis of human personality and robot acceptance in human-robot interaction. Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021. [doi: 10.1145/3411764.3445542]
52. Chatzoglou PD, Lazaraki V, Apostolidis SD, Gasteratos AC. Factors affecting acceptance of social robots among prospective users. Int J of Soc Robotics 2024 Jun;16(6):1361-1380. [doi: 10.1007/s12369-023-01024-x]
53. The Criteria Committee of the New York Heart Association. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels, 9th edition: Little Brown & Co; 1994:253-256.
54. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). J Med Libr Assoc 2018;106(1):142-144. [doi: 10.5195/jmla.2018.319]
55. Deutsch I, Erel H, Paz M, Hoffman G, Zuckerman O. Home robotic devices for older adults: opportunities and concerns. Comput Human Behav 2019 Sep;98:122-133. [doi: 10.1016/j.chb.2019.04.002]
56. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
57. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
58. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
59. Hammash MH, Hall LA, Lennie TA, et al. Psychometrics of the PHQ-9 as a measure of depressive symptoms in patients with heart failure. Eur J Cardiovasc Nurs 2013 Oct;12(5):446-453. [doi: 10.1177/1474515112468068] [Medline: 23263270]
60. van der Zwaan GL, van Dijk SEM, Adriaanse MC, et al. Diagnostic accuracy of the Patient Health Questionnaire-9 for assessment of depression in type II diabetes mellitus and/or coronary heart disease in primary care. J Affect Disord 2016 Jan 15;190:68-74. [doi: 10.1016/j.jad.2015.09.045] [Medline: 26480213]
61. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
62. Bolgeo T, Di Matteo R, Simonelli N, et al. Psychometric properties and measurement invariance of the 7-item General Anxiety Disorder scale (GAD-7) in an Italian coronary heart disease population. J Affect Disord 2023 Aug 1;334:213-219. [doi: 10.1016/j.jad.2023.04.140] [Medline: 37149049]
63. Su R, Tay L, Diener E. The development and validation of the Comprehensive Inventory of Thriving (CIT) and the Brief Inventory of Thriving (BIT). Appl Psychol Health Well Being 2014 Nov;6(3):251-279. [doi: 10.1111/aphw.12027] [Medline: 24919454]
64. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the big-five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
65. Thørrisen MM, Sadeghi T. The Ten-Item Personality Inventory (TIPI): a scoping review of versions, translations and psychometric properties. Front Psychol 2023;14:1202953. [doi: 10.3389/fpsyg.2023.1202953] [Medline: 37434881]
66. Adachi T, Tsunekawa Y, Matsuoka A, Tanimura D. Association between big five personality traits and participation in cardiac rehabilitation in Japanese patients with cardiovascular disease: a retrospective cohort study. Int J Environ Res Public Health 2021 Aug 14;18(16):8589. [doi: 10.3390/ijerph18168589] [Medline: 34444339]
67. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. Management Information Systems Quarterly 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]

XSL•FO
RenderX

68. Venkatesh V, Thong J, Xu X, University of Arkansas, Hong Kong University of Science and Technology, The Hong Kong Polytechnic University. Unified Theory of Acceptance and Use of Technology: a synthesis and the road ahead. JAIS 2016;17(5):328-376. [doi: 10.17705/1jais.00428]
69. David D, Thérouanne P, Milhabet I. The acceptability of social robots: a scoping review of the recent literature. Comput Human Behav 2022 Dec;137:107419. [doi: 10.1016/j.chb.2022.107419]
70. Philippi P, Baumeister H, Apolinário-Hagen J, et al. Acceptance towards digital health interventions - model validation and further development of the unified theory of acceptance and use of technology. Internet Interv 2021 Dec;26:100459. [doi: 10.1016/j.invent.2021.100459] [Medline: 34603973]
71. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45(3):1-68. [doi: 10.18637/jss.v045.i03]
72. Hemphill JF. Interpreting the magnitudes of correlation coefficients. Am Psychol 2003;58(1):78-79. [doi: 10.1037/0003-066X.58.1.78]
73. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Taylor & Francis; 2013.
74. Aberson CL. Applied Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2019.
75. Kessing D, Denollet J, Widdershoven J, Kupper N. Psychological determinants of heart failure self-care: systematic review and meta-analysis. Psychosom Med 2016 May;78(4):412-431. [doi: 10.1097/PSY.0000000000000270] [Medline: 27082055]
76. Heerink M, Kröse B, Wielinga B, Evers V. Enjoyment intention to use and actual use of a conversational robot by elderly people. Presented at: HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction; Mar 12-15, 2008; Amsterdam The Netherlands p. 113-120. [doi: 10.1145/1349822.1349838]
77. Goodlin SJ, Gottlieb SH. Social isolation and loneliness in heart failure: integrating social care into cardiac care. JACC Heart Fail 2023 Mar;11(3):345-346. [doi: 10.1016/j.jchf.2023.01.002] [Medline: 36889882]
78. Ghafurian M, Ellard C, Dautenhahn K, et al. Social companion robots to reduce isolation: a perception change due to COVID-19. In: Ardito C, Lanzilotti R, Malizia A, Petrie H, Piccinno A, Desolda G, editors. Human-Computer Interaction – INTERACT 2021: Springer International Publishing; 2021:43-63. [doi: 10.1007/978-3-030-85616-8_4]
79. Kim S, Zhong Y, Wang J, Kim HS. Exploring technology acceptance of healthcare devices: the moderating role of device type and generation. Sensors (Basel) 2024;24(24):7921. [doi: 10.3390/s24247921]
80. Park J, Lee J, Noh D. Mobile app for improving the mental health of youth in out-of-home care: development study using an intervention mapping approach. JMIR Hum Factors 2024 Nov 21;11:e64681. [doi: 10.2196/64681] [Medline: 39571152]
81. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
82. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106. [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
83. Guemghar I, Pires de Oliveira Padilha P, Abdel-Baki A, Jutras-Aswad D, Paquette J, Pomey MP. Social robot interventions in mental health care and their outcomes, barriers, and facilitators: scoping review. JMIR Ment Health 2022 Apr 19;9(4):e36094. [doi: 10.2196/36094] [Medline: 35438639]
84. Ke C, Lou VWQ, Tan KCK, Wai MY, Chan LL. Changes in technology acceptance among older people with dementia: the role of social robot engagement. Int J Med Inform 2020 Sep;141:104241. [doi: 10.1016/j.ijmedinf.2020.104241]
85. Henschel A, Laban G, Cross ES. What makes a robot social? A review of social robots from science fiction to a home or hospital near you. Curr Robot Rep 2021;2(1):9-19. [doi: 10.1007/s43154-020-00035-0] [Medline: 34977592]
86. Lund LH, Crespo‐Leiro MG, Laroche C, et al. Heart failure in Europe: guideline‐directed medical therapy use and decision making in chronic and acute, pre‐existing and de novo, heart failure with reduced, mildly reduced, and preserved ejection fraction – the ESC EORP heart failure III registry. European J of Heart Fail 2024 Dec;26(12):2487-2501. [doi: 10.1002/ejhf.3445]
87. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: a meta-analysis. Computers & Education 2017 Feb;105:1-13. [doi: 10.1016/j.compedu.2016.11.003]
88. He Y, He Q, Liu Q. Technology acceptance in socially assistive robots: scoping review of models, measurement, and influencing factors. J Healthc Eng 2022;2022:6334732. [doi: 10.1155/2022/6334732] [Medline: 35911583]
89. Jones NL, Loizides F, Jones K. Acceptance of needy socially assistive robots: a systematic review. Int J Soc Robot 2025 Jul;17(7):1289-1303. [doi: 10.1007/s12369-025-01272-z]
90. Jan J, Alshare KA, Lane PL. Hofstede's cultural dimensions in technology acceptance models: a meta-analysis. Univ Access Inf Soc 2024 Jun;23(2):717-741. [doi: 10.1007/s10209-022-00930-7]
91. Prichard R, Maneze D, Straiton N, Inglis SC, McDonagh J. Strategies for improving diversity, equity, and inclusion in cardiovascular research: a primer. Eur J Cardiovasc Nurs 2024 Apr 12;23(3):313-322. [doi: 10.1093/eurjcn/zvae002] [Medline: 38190724]

Abbreviations

BIT: Brief Inventory of Thriving
CCI: Charlson Comorbidity Index

XSL•FO
RenderX

GAD-7: Generalized Anxiety Disorder-7 scale
HF: heart failure
HFrEF: Heart failure with reduced ejection fraction
LVEF: left ventricular ejection fraction
NYHA: New York Heart Association
PHQ-9: Patient Health Questionnaire-9
PPWB: positive psychological well-being
REDCap: Research Electronic Data Capture
SR: social robot
TIPI: Ten-Item Personality Inventory
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Coristine; submitted 28.Aug.2025; peer-reviewed by N Neuendorf, U Pasupulety; revised version received 07.Apr.2026; accepted 07.Apr.2026; published 02.Jun.2026.

Please cite as:
Maukel LM, Bouchard K, Fiedorowicz JG, Dautenhahn K, Ghafurian M, Coutinho T, McGuinty C, Liu PP, Tulloch H
Positive Psychological Well-Being and Determinants of Social Robot Acceptability Among Patients With Heart Failure: Cross-Sectional Questionnaire Study
JMIR Cardio 2026;10:e83163
URL: https://cardio.jmir.org/2026/1/e83163
doi:10.2196/83163

© Lisa-Marie Maukel, Karen Bouchard, Jess G. Fiedorowicz, Kerstin Dautenhahn, Moojan Ghafurian, Thais Coutinho, Caroline McGuinty, Peter P. Liu, Heather Tulloch. Originally published in JMIR Cardio (https://cardio.jmir.org), 2.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Massive Open Online Course for Primary Prevention of Stroke: Evaluation of Co-Design Process

Aastha Gurung[1], BPsychSc, BHSc (Hons); Seamus Barker[1], BPT, BA (Hons), MPhil, PhD; Sabah Rehman[2], MBBS, PhD; Rachel E Climie[1], BExSc (Hons), PhD; Rosanne Freak-Poli[3,4], BHS, PhD; Catherine Burns[3], BSc, MHS, MHlthInfoMgmt; Monique F Kilkenny[3,5], MPH, PhD; Tara Purvis[3], BPhysio (Hons), MSc, PhD; Christine Farmer[1], BSc, GradDipMarAntSc, MEnv; Brenda Booth[6], RN; Janet E Bray[4], PhD; Dominique A Cadilhac[3,5], MPH, PhD; Jan Cameron[3], BN, GradDipHlthSc, MHS, PhD; Lachlan L Dalli[3], BBiomedSc (Hons), PhD; Stephanie Ho[6], BBus; Eleanor Horton[6], PhD; Timothy Kleinig[7], MBBS (Hons), BA, PhD; Lisa Murphy[6], BSc, MBBS, MBA; Mark R Nelson[1], MBBS (Hons), MFM, PhD; Muideen T Olaiya[3], BSc (Hons), MSc, PhD; Amanda G Thrift[3], PhD; Seana L Gall[1], BSc (Hons), PhD

[1]Menzies Institute for Medical Research, University of Tasmania, 17 Liverpool St, Hobart, Tasmania, Australia

[2]Tasmanian School of Medicine, University of Tasmania, Hobart, Tasmania, Australia

[3]Stroke and Ageing Research, Department of Medicine, School of Clinical Sciences, Monash University, Clayton, Victoria, Australia

[4]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

[5]Stroke and Critical Care Research Theme, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia

[6]Stroke Foundation, Melbourne, VIC, Australia

[7]Adelaide Medical School, The University of Adelaide, Adelaide, South Australia, Australia

Corresponding Author:
Seana L Gall, BSc (Hons), PhD
Menzies Institute for Medical Research, University of Tasmania, 17 Liverpool St, Hobart, Tasmania, Australia

Abstract

Background: Public awareness and understanding of stroke are crucial to reducing its overall burden. Massive open online courses (MOOCs) are free, self-directed online learning platforms that provide accessible education.

Objective: This study aimed to describe the process, outcomes, and evaluation of co-designing the Preventing Stroke MOOC.

Methods: The co-design process was conducted with various stakeholder groups from July 2020 to December 2024 in Australia over three phases: (1) predesign, (2) co-design, and (3) postdesign. In the predesign phase, we consulted with MOOC researchers and held a half-day in-person workshop with community members to prioritize the stroke prevention topics, structure, format, and features, followed by user testing of a single developed module. The larger study comprised 16 virtual focus groups (14 co-design and 2 user testing focus groups) with community members and health knowledge experts. Of the 14 co-design focus groups, 12 informed the MOOC design, including content, structure, and delivery. This was followed by drafting content for each module, expert review (eg, clinicians and researchers), and development of platform materials. In the postdesign phase, we conducted user testing to collect feedback and evaluated the co-design process to determine how participants' input shaped the final MOOC design.

Results: In the predesign phase, 5 MOOC experts and community members contributed to a pilot module comprising 9 pages on the prioritized topics, including a focus on stories from people with lived experience of stroke. A total of 8 people tested this module, providing positive feedback about content and delivery. The co-design phase engaged 22 participants across 12 focus groups relevant to the MOOC. Their suggestions contributed to the multimedia design of the 7 core modules, with personalization via 9 optional mini-modules. Feedback on technical issues and suggested improvements identified during postdesign user testing were resolved. Evaluation of the MOOC co-design process found that the majority of participants' suggestions gathered during the co-design phase were included in the final MOOC.

Conclusions: A structured, co-design process resulted in the Preventing Stroke MOOC reflecting participants' suggestions and provided important insights on the educational information needed to inform the general population about stroke prevention.

(JMIR Cardio 2026;10:e89358) doi:10.2196/89358

KEYWORDS

stroke; cardiovascular disease; co-design; coproduction; cocreation; massive open online course; preventive medicine; MOOC

XSL•FO
RenderX

Introduction

Stroke is largely preventable, with a set of common modifiable risk factors accounting for nearly 90% of the population-attributable risk of stroke [1]. Modifiable risk factors include physical inactivity, smoking, inadequate diet, and excessive alcohol consumption, as well as biomedical factors such as hypertension, atrial fibrillation, diabetes, obesity, and dyslipidemia [1,2]. Such factors are common, with almost all adults in Australia having at least 1 risk factor and nearly 88% having 2 or more risk factors for stroke [3]. Consequently, increasing public knowledge of risk factors for stroke is crucial for enhancing prevention efforts [4-6]. In Australia, the Stroke Foundation runs 2 educational campaigns aimed at improving the knowledge of stroke risk factors and warning signs in the community [7-9]. The first campaign involves mass media promotion of face, arms, speech, and time (FAST) stroke signs. The second campaign is the StrokeSafe Speaker program, which involves presentations to the community, usually delivered by volunteers with lived experience of stroke. While there has been an improvement in the understanding of stroke in the community since the first national campaign in 2004 [7-9], many Australians cannot identify stroke warning signs and fail to call emergency medical services when a stroke occurs [10]. This may be related to the incomplete retention of knowledge and limited behavior change from these types of educational interventions [6,7]. Kilkenny et al [7] demonstrated that the knowledge gained by attendees to the Stroke Foundation's StrokeSafe Speaker program diminishes within 3 months posteducation. An important consideration in educating the community about stroke is that interventions should achieve sustained knowledge and actions toward stroke prevention.

The global growth of internet use has created opportunities to provide accessible health information, interventions, and services [11]. Massive open online courses (MOOCs) are free and flexible web-based learning tools [12,13]. MOOCs enable the delivery of education to a large audience through video lectures, quizzes, additional resources, and discussion boards [12]. Numerous MOOCs are available covering a wide range of health promotion topics [14]. A search of major online learning platforms, including MOOC List, edX, and Coursera, using the keyword "stroke" identified courses that covered topics such as stroke anatomy, recovery, and endovascular treatment patient selection, but none specifically addressed stroke prevention. To the best of our knowledge, there are currently no MOOCs focused on primary stroke prevention. Such a MOOC has the potential to educate the general public with evidence-based information about stroke, increase awareness of prevention, support recognition of individual risk, and promote positive behavior change.

Participatory research approaches, such as co-design, are widely used to develop inclusive health interventions [15]. Co-design is an iterative process involving active and meaningful collaboration between stakeholders with diverse expertise throughout the design process [16], with the aim of tailoring the intervention to the needs of the intended audience [15-17]. Co-design has been increasingly used to develop effective and sustainable health interventions [18]. For MOOCs aimed at improving health-related knowledge and understanding, the quality of educational content plays a critical role in achieving learning outcomes [19,20]. Co-designing MOOCs with stakeholders, informed by real-world needs and experiences, can enhance content relevance for intended users [19].

While co-design approaches are widely used in developing health interventions, limited emphasis has been placed on evaluating the co-design process [21]. Evaluating the co-design process is essential to determine whether the contributions of co-design participants, as active partners, were carried through to the end product, to strengthen understanding of how the process influenced intervention development, and to identify areas for improvement in future co-design activities [21].

Our aim was to undertake a co-design approach to develop a MOOC for primary stroke prevention to address poor knowledge and improve the identification and management of stroke risk factors in the community. Specifically, we aimed to:

- Describe the process and outcomes of co-designing the Preventing Stroke MOOC.
- Evaluate the co-design process of the Preventing Stroke MOOC to determine the extent to which participants' suggestions from the co-design sessions were included in the final MOOC design.

Methods

Study Design

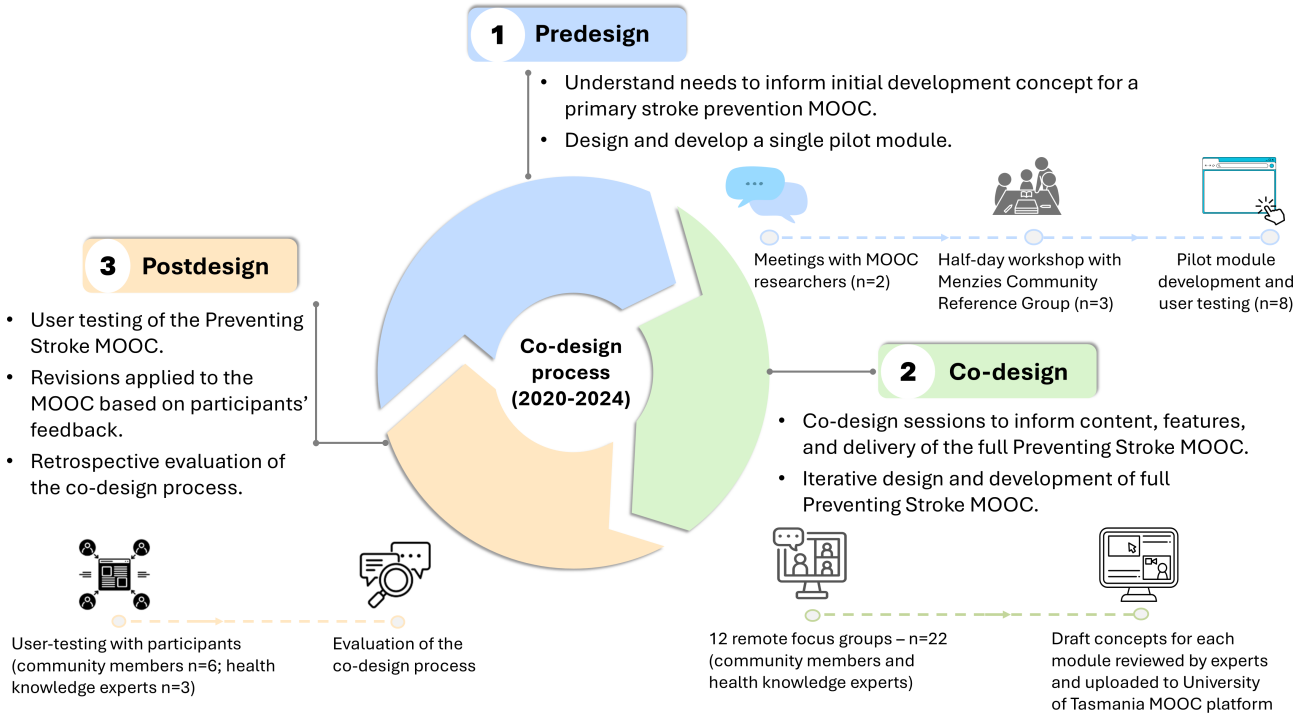
This study is part of a larger research project named "Love Your Brain: A digital health intervention to prevent stroke," comprising two components: (1) a MOOC and (2) a text messaging system [22]. The co-design method [23], overall co-design results across both components of the platform [24], and the development of the text message system [25] have been described previously. This study used a qualitative research approach to co-design the MOOC and retrospectively evaluate the co-design process. Methods and results are reported according to the Consolidated criteria for Reporting Qualitative research (COREQ) checklist (Checklist 1) [26].

Overview of Co-Design Process

The co-design of the Preventing Stroke MOOC was carried out in three iterative phases involving (1) predesign, (2) co-design, and (3) postdesign (Figure 1), broadly aligned with the generative co-design framework by Bird et al [27]. The co-design process occurred between July 2020 and December 2024. Data collection included consultation meetings, workshops, and focus groups, which were audio and video recorded where appropriate and supplemented by field notes. Participants in the co-design process comprised diverse stakeholder groups, including community members (with or without a history of stroke) and health knowledge experts (eg, allied health clinicians, nurses, and researchers). Throughout the process, participants were active partners, positioned as experts of their lived experience.

XSL•FO
RenderX

Figure 1. Co-design process. MOOC: massive open online course.



Phase 1: Predesign (July 2020-March 2022)

Collecting Information to Understand Needs

The first stage of the predesign phase involved consultation meetings with researchers at the University of Tasmania (UTAS) who have experience in MOOC design and development. The aim was to understand the requirements related to the initial development concept for a stroke prevention MOOC. These researchers were emailed an invitation to participate in informal in-person consultation meetings on July 20 and August 13, 2020. The purpose of these meetings was to gain insights into their experience in designing health MOOCs for the community.

We then explored community perspectives on how a MOOC could support primary stroke prevention. Participants were recruited via convenience sampling of Menzies Institute for Medical Research (MIMR) Cardiovascular Research Flagship Community Reference Group (CRG). CRG members were invited via email to participate in an in-person half-day workshop at MIMR in March 2022. The workshop was facilitated by the MIMR Cardiovascular Research Flagship lead (SLG, female, PhD), a cardiovascular epidemiologist with experience in conducting qualitative research. The workshop focused on the perceived usefulness of a stroke prevention MOOC, topics for inclusion, preferred delivery modes, and strategies to engage learners. The Cardiovascular Research Flagship program manager (CF, female, Master of Environment [MEnv]) synthesized findings from this stage to consolidate insights from the meeting and workshop, guiding the initial design concept for the MOOC. This involved reading and interpreting the field notes from both sessions.

Development and User Testing of the Pilot Module

Based on the needs identified during the first stage, we developed a single pilot module to present the initial concept for the proposed Preventing Stroke MOOC. The module was drafted using existing evidence and resources, with content modeled from the Stroke Foundation's StrokeSafe Speaker program.

Materials for the module included videos featuring experts (eg, epidemiologists and cardiologists) and people with lived experience (eg, survivors of stroke), who were invited to share their insights and stories. Experts from the research team's network were invited to participate in video recordings, where they discussed stroke and risk factors from clinical, research, and public health perspectives. Invitations were sent to CRG members with lived experience of stroke to participate in video recordings to share their experiences of living with and managing stroke. Each video participant drafted their own script, which the research team then reviewed and refined together with them as needed. The filming took place in person in Hobart, Tasmania. The final draft of the module and video transcripts were reviewed by other members of the local research team and refined through iterative revisions until consensus was reached. The pilot module was developed in collaboration with the UTAS Digital Futures team, who supported page design, video uploads, and graphics integration. The pilot module was uploaded to the UTAS MOOC platform with restricted access.

Snowballing methods were used to recruit participants for user testing. Email invitations were circulated via the Stroke Foundation Lived Experience Council (formerly the Consumer Council). The Lived Experience Council is a board subcommittee within Stroke Foundation that provides insights into strategy and operations. Participants who agreed to take part were emailed a link to access the module. While working through the module, they were asked to reflect on the usefulness of topics, aspects they found engaging, any technical difficulties, and suggestions for informing future modules. To obtain feedback on the pilot module, we conducted 3 focus groups

XSL•FO
RenderX

between March 3 and 23, 2022. The focus groups followed a semistructured guide (Multimedia Appendix 1). All sessions were facilitated by the flagship program manager (CF) and attended by the flagship lead (SLG), who took notes and responded to participant queries as needed. One researcher (AG) reviewed focus group transcripts and field notes to summarize participants' feedback on the pilot module.

Phase 2: Co-Design (April 2023-February 2024)

Data Collection Through Co-Design Focus Groups

The Preventing Stroke MOOC was part of the Love Your Brain project, which also involved co-designing a text message system [24]. Participants included two separate groups: (1) community members and (2) health knowledge experts involved in health education or promotion. Recruitment occurred between April and May 2023. Community members were approached using recruitment flyers distributed by email to consumer reference groups (eg, MIMR CRG and Stroke Foundation) and newsletters (eg, Heart Foundation consumer newsletter). Health knowledge experts were recruited through emails circulated within the research team's professional networks, including research collaborators and clinicians (eg, nurses, stroke neurologists, physiotherapists, and occupational therapists). Focus group methods, including facilitation approach, discussion guide, and analysis, are described in detail elsewhere and are freely accessible [24]. A total of 16 focus groups were conducted as part of the larger study (14 co-design focus groups and 2 user-testing focus groups), of which 12 co-design focus groups were relevant to the MOOC and are reported in this phase [24]. In brief, separate sessions were held remotely via Zoom (Zoom Communications, Inc) for each participant group to mitigate power imbalances. Discussion topics included prompt questions framed around primary stroke prevention (see Multimedia Appendix 2 for the question list used in each focus group). Over 4 months (June-September 2023), participants worked together with the research team to inform design decisions for the MOOC, including content, level of detail, presentation format, delivery modes, and engagement strategies [24].

Development of Preventing Stroke MOOC

Findings from the co-design focus groups, existing guidelines on primary stroke prevention, and evidence from the literature informed the development of the MOOC. Video materials were developed for the modules, incorporating lived experience stories and insights from subject area experts. Some community members who participated in the co-design focus groups were invited to take part in filming. Subject area experts (eg, epidemiologists, cardiologists, neurologists, and paramedics) were recruited through the research team's professional networks. Video scripts were drafted by speakers and reviewed by the research team before final review and approval by the speaker. Overall, 2 types of video styles were used based on findings from the co-design process, including direct-to-camera recordings and hand-drawn animations. Direct-to-camera videos featured speakers delivering short, scripted segments on topics relevant to their expertise. Animated videos were produced in collaboration with Russ Animates Ltd. A transcript of each video was available to download as a PDF to ensure accessibility.

Interactive features, including gamified activities such as flashcards and crosswords, were developed to ensure the content was engaging. End-of-module quizzes were created for each module, and links to additional resources were selected by the research team and subject area experts to complement the topics. Once the content and materials were finalized, they underwent iterative review and refinement in collaboration with subject area experts. The Digital Futures team of education technologists at UTAS supported the design of the MOOC platform. The content was uploaded to the MOOC page hosted by UTAS with restricted access. The entire development process took place over a 9-month period (May 2023-February 2024).

Phase 3: Postdesign (March 2024-December 2024)

Preventing Stroke MOOC User Testing Data Collection

Invitations to participate in user testing of the MOOC were sent to the same participants who took part in the focus groups during the co-design phase. Additional participants were invited through the research team's professional networks to provide broader perspectives. Participants who agreed to participate in user testing received a welcome email containing a link to access the MOOC and were given 8 weeks to engage with the course.

Feedback was collected through separate focus groups with the community and health expert groups remotely via Zoom in March 2024. The data collection method followed the same process as that used for the co-design focus groups [24]. The sessions focused on exploring users' experiences with the MOOC, including overall perceptions of the course, features they found useful, challenges encountered, and suggestions for improvement. In addition, 3 health knowledge experts who had not participated in the co-design phase were provided access to the MOOC and asked to review the course and provide written feedback via email regarding content, usability, and suggestions for improvement. The research team reviewed the feedback and identified areas for improvement. These were discussed in team meetings and shared with the UTAS Digital Futures team, with revisions made to the MOOC where feasible, technically possible, and within budget.

Preventing Stroke MOOC User Testing Data Analysis

Transcripts from the user testing focus group were analyzed manually in Microsoft Word. One researcher (AG) reviewed and summarized participants' feedback using a deductive approach [28]. Feedback from the community member focus group and email feedback from health knowledge experts were organized according to the main categories of the user-testing discussion guide (eg, general impressions, useful features, challenges encountered, and suggestions for improvement) and grouped under a single overarching theme reflecting overall feedback on the MOOC. The process involved frequent discussion with one of the focus group facilitators (SB, male, Research Fellow, PhD), and findings were reviewed and confirmed by the research team.

Evaluation of the Co-Design Process

We conducted qualitative content analysis to retrospectively evaluate the co-design process and determine whether participants' input from the co-design phase was incorporated

XSL•FO
RenderX

into the final MOOC design [28,29]. The analysis followed 3 steps.

In the first step, one researcher (AG) engaged with the MOOC platform as a mock participant and worked through each module. The MOOC was treated as qualitative data, and all module content was reviewed in detail. As each module was completed, key design features of the course were identified and described (eg, whether learning objectives were present, the structure of modules, topics covered, the number and type of learning activities, and the media formats used). These descriptions were used to characterize how the MOOC was designed. Word counts were calculated for each module, including video transcripts, using Microsoft Word, and readability was assessed using the Flesch Reading Ease score [30].

In the second step, 12 transcripts from co-design focus groups involving 22 participants were imported into the NVivo software (version 15; Lumivero). The analysis began with a close reading of the transcripts. Meaning units were identified inductively as words, sentences, or short passages conveying a single idea related to the MOOC [28,29]. These meaning units were then interpreted to understand participants' suggestions in relation to the MOOC design. An initial subset of transcripts was analyzed first to explore how participants' suggestions were being interpreted. These interpretations were discussed with 2 researchers (SB and SLG), who were involved in the co-design facilitation, to ensure the analysis was credible and meaningful within the context of the co-design process. The remaining transcripts were then analyzed, with ongoing discussion with SB and SLG. This involved revisiting selected excerpts, considering alternative interpretations, and refining the understanding of participants' suggestions.

In the final step, we undertook triangulation by comparing findings from the MOOC and the co-design focus groups to examine whether and how participants' suggestions were reflected in the final MOOC design [31]. To support this comparison, we drew on design elements adapted from the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) instructional design model [32], which outlines key aspects of course design, comprising learning objectives, structure, content, exercise, assessment, and media selection. Features identified in the MOOC were organized in relation to these design elements (Multimedia Appendix 3), and participants' suggestions were subsequently examined alongside corresponding MOOC features for each design element. The number of meaning units identified in the second step was summarized descriptively by design element during this process to support transparency in reporting the distribution of participants' suggestions. An example of the analytic process is provided in Multimedia Appendix 4.

Trustworthiness of the analysis was supported through strategies aligned with criteria for qualitative content analysis, including credibility, dependability, and confirmability [33]. Credibility was supported through ongoing discussions of interpretations with researchers who were involved in the co-design facilitation and contributed to interpreting the data. Dependability was supported through clear documentation of the data analysis process. Confirmability was supported by revisiting the transcripts and discussing interpretations to ensure they were grounded in the data. The analysis was primarily conducted by a researcher (AG) who was not involved in the co-design process, which helped reduce the influence of prior assumptions when interpreting participants' suggestions and how they were reflected in the MOOC. In this context, interrater reliability was not considered appropriate for the approach taken, as the aim was interpretive understanding, with interpretations discussed with the research team [34,35].

Ethical Considerations

Ethical approval for this study was obtained from the UTAS Human Research Ethics Committee (#27108) and the Monash University Human Research Ethics Committee (#35899; Universal Trial Number U1111-1305-2964). Verbal informed consent was obtained from all participants prior to participation, including consent for audio and video recording. This approach to consent was reviewed and approved by the human research ethics committees of each institution. Participants were reimbursed Aus $50 (US $34.70) for their time. Reimbursement was provided via bank transfer or gift voucher, according to participant preference. Participants were assigned unique study identification numbers, and identifying information was removed from transcripts prior to analysis. Study data were stored securely and were accessible only to authorized members of the research team.

Results

Phase 1: Predesign (July 2020-March 2022)

Collecting Information to Understand Needs

Two researchers involved in the design and development of the Preventing Dementia MOOC and the Understanding Multiple Sclerosis MOOC attended separate consultation meetings with the research team. Together, they recommended the following to guide the development of a successful MOOC: (1) engage stakeholders early in the design process, (2) plan for development costs, and (3) provide feedback to learners.

A total of 3 CRG members participated in the half-day workshop. They emphasized the importance of personalizing content to make it relevant and engaging for learners. Key elements recommended for inclusion were personalized stroke risk assessments and interactive features such as quizzes and discussion boards. Participants highlighted that the MOOC content should follow a structure similar to the StrokeSafe Speaker program, feature stories from people with lived experience of stroke, and include multimodal delivery through a combination of videos, graphics, and text to accommodate different learning styles.

Development and User Testing of a Pilot Module

The pilot module comprised 9 pages and covered topics that broadly followed the StrokeSafe Speaker program, including an introduction to stroke, living with stroke, stroke as a medical emergency, stroke statistics, understanding stroke risk, and managing stroke risk. The module featured 4 videos, with 2 presented by experts (a cardiovascular epidemiologist and a general practitioner) and 2 by people with lived experience of

XSL•FO
RenderX

stroke. Additional materials included an end-of-module quiz, a discussion board, links to further resources, and a glossary.

A total of 9 participants were recruited for user testing of the pilot module. One participant opted out due to COVID-19. Participants were from Tasmania, New South Wales, Victoria, and Western Australia. Overall, 2 focus groups were held remotely via Zoom (n=6 participants), and 1 focus group was conducted in person at MIMR (n=2 participants). Each session lasted approximately 30 to 60 minutes.

All participants reported using a laptop or desktop computer to complete the pilot module. Overall, feedback was positive regarding the structure of the content, the topics covered, and the materials included. Participants described the module as engaging and easy to follow, indicating that it could be a useful tool for the general public. They further expressed that videos from people with lived experience of stroke had a particularly strong impact. Participants suggested that the full MOOC should include subtitles in the videos to improve accessibility. They also suggested that additional optional mini-modules could be incorporated, such as dedicated sections covering aphasia and childhood stroke.

Phase 2: Co-Design (April 2023-February 2024)

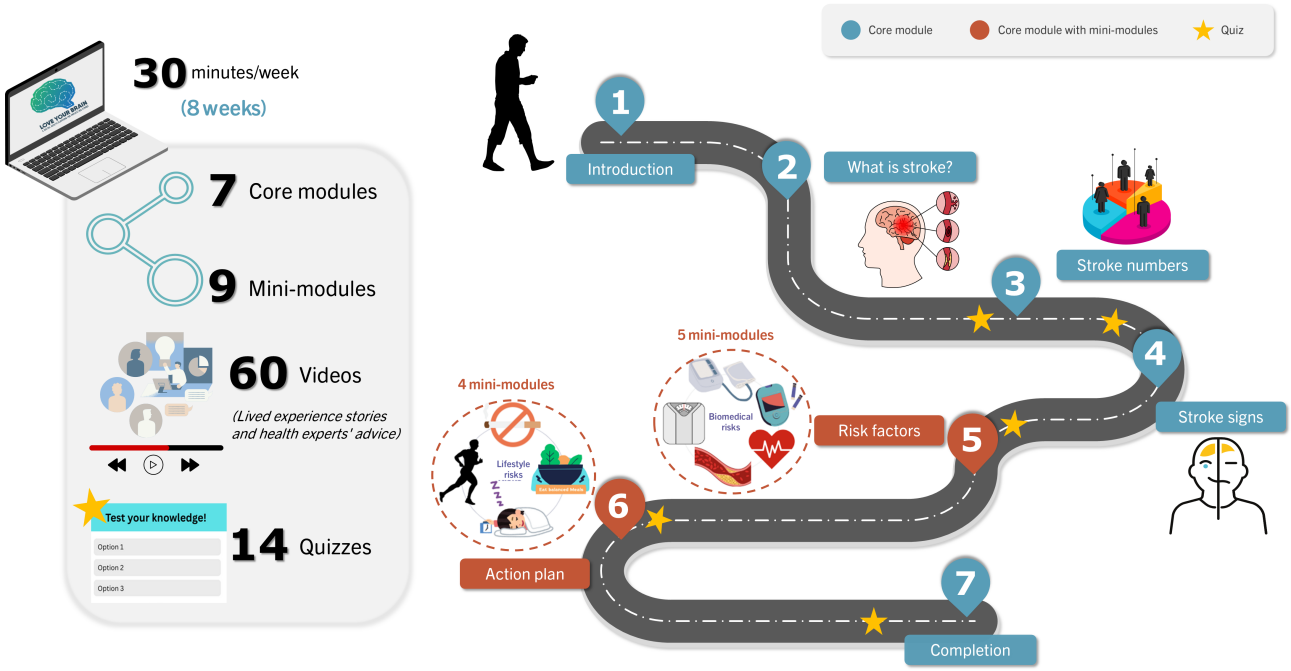
Co-Design Focus Groups

In total, 22 purposively selected participants participated in at least 1 focus group (n=12 community members and n=10 health knowledge experts). Among the community members participating, 58% (7/12) were female, 58% (7/12) were aged <65 years, 75% (9/12) were stroke survivors, and 25% (3/12) had no personal history of stroke. Health knowledge experts were predominantly female (8/10, 80%), aged >35 years (6/10, 60%), and comprised equal proportions of researchers (5/10, 50%) and clinicians (5/10, 50%). Comprehensive thematic analysis of the co-design focus groups can be found in our previous publication [24]. In summary, specific to the MOOC, participants in the focus groups suggested including topics on stroke signs, risk factors, and the impact of stroke [24]. They emphasized ensuring the accuracy and credibility of information, the use of plain language, multimedia delivery formats, personalization through self-paced modules and topic choice based on learners' risk factors, incorporating interactive elements, stories from people with lived experience of stroke, accessibility across major devices, and inclusive design [24]. They also specifically mentioned the benefit in using well-recognized evidence-based resources, including topics from the StrokeSafe Speaker program, to guide the structure of the MOOC content [24].

Development of Preventing Stroke MOOC

In total, 16 modules were designed, with 7 core modules and 9 optional mini-modules (Figure 2). The seven core content modules were (1) Introduction, (2) What is stroke, (3) Stroke numbers, (4) Stroke signs, (5) Risk factors for stroke, (6) Action plan, and (7) Completion. Optional mini-modules were designed to support personalization. These included 5 mini-modules focused on biomedical risk factors (blood pressure, cholesterol, atrial fibrillation, overweight and obesity, and blood sugar) and 4 mini-modules focused on lifestyle-related risk factors (smoking, diet and alcohol, exercise, and sleep). Participants were encouraged to complete at least 1 mini-module from each section. Participants received a certificate in the final module once they completed the course. A total of 60 individual videos were filmed with subject area experts (n=13) and individuals with lived experience of stroke or carers (n=8). We also developed 6 interactive activities such as flashcards and crosswords, 14 end-of-module quizzes, 110 graphics or infographics, and 129 hyperlinks to additional resources integrated into relevant modules.

Figure 2. Preventing Stroke Massive Open Online Course learner journey.



XSL•FO
RenderX

Phase 3: Postdesign (March 2024-December 2024)

Preventing Stroke MOOC User Testing

A total of 6 community members who had participated in the co-design focus groups undertook user testing of the MOOC. Health knowledge experts who participated in the co-design focus groups did not respond to the MOOC user testing invitation. Additional feedback was obtained via email from 3 health knowledge experts who had not been involved in the co-design phase. The focus group with community members was conducted online on March 13, 2024, and lasted approximately 60 minutes.

Overarching feedback from participants indicated that the MOOC was user-friendly, engaging, and well-structured. Participants also reported that the interactive features were useful and that the personalized risk assessment and videos provided impact. Participants noted a few technical issues, including difficulties accessing the MOOC via mobile phones and tablets.

Suggestions for improvement included defining key terms, maintaining consistent terminology throughout the content, and providing clear signposting, such as bolding keywords to support understanding and comprehension. Two participants noted that, from a disability perspective, graphics using similar colors may pose challenges for individuals with vision impairment or color blindness. They recommended high-contrast colors to improve visual accessibility. In collaboration with the Digital Futures team at UTAS, strategic modifications were made. These included technical updates to support compatibility across all major devices, refinement of wording for consistency, use of bold text for keywords, clear signposting, and use of high-contrast colors in graphics.

Evaluation of the Co-Design Process

Following review and documentation of the MOOC, the content covered a total of 81 topics, and multiple-choice questions (MCQs) were included in 14 modules (Multimedia Appendix 3). Data from co-design focus group transcripts (step 2) were analyzed to identify participants' suggestions related to MOOC design, with 223 meaning units identified across the 12 transcripts. Triangulation of findings from step 1 and 2 involved examining how these suggestions were reflected in the MOOC. This comparison was conducted across the adapted design elements, with illustrative quotes presented in Tables 1-4. The findings are presented below, organized according to these elements.

Table . Participant suggestions from co-design focus groups (step 2) and comparison with MOOC[a] design (step 3) for learning objectives and structure.

Design elements	Step 2: co-design focus groups			Step 3: triangulation	
	Participants' suggestion	Meaning units	Example quotes	Suggestions integrated into MOOC	Observations
Learning objectives	• Clear guidelines and expectations	2	"Having all those really clear set guidelines at the beginning would be an incentive, to know whether the people are going to sign up and keep going." (Community member, P2)	Clearly stated learning objectives were present in all core modules, except for Module 7, which was a completion module and therefore did not require learning objectives.	Inconsistency in learning objectives across mini modules. 4 out of 9 mini modules did not state learning objectives
Structure	• Layer knowledge across modules to build understanding • Core modules for foundational learning, with optional modules on specific topics	11	"I thought it would be good to encapsulate all the information in a sequence…" (Community member, P3) "I really love the idea of having kind of like a sliding scale of the depth of information. So, you could have basic, moderate, and hardcore. Because I feel like even the general public sort of have this sliding scale of wanting information. Some people are happy with the basics, some people are thirsty for more." (Health knowledge expert, P2)	Structure of almost all modules was linear. Module 5 and Module 6 core module which included mini modules had modular learning. structure	Suggestions were addressed.

[a]MOOC: massive open online course.

XSL•FO
RenderX

Table . Participant suggestions from co-design focus groups (step 2) and comparison with MOOC[a] design (step 3) for content.

Design elements	Step 2: co-design focus groups			Step 3: triangulation	
	Participants' suggestion	Meaning units	Example quotes	Suggestions integrated into MOOC	Observations
Content	• Explain the word 'stroke,' use anatomical references • Emphasize that stroke can affect people of all ages • Include statistics • Importance of recognizing the FAST[b] signs of stroke • Address modifiable and nonmodifiable risk factors • Visible and invisible impact of stroke • Content should be simple and easy to understand • Provide checklists and action plans for health appointments • Include personal stories to enhance relatability	127	• "Empowering of people to take those steps, to actually go to the doctor and to - I think prevention is so much about that empowerment and agency of people…" (Community member, P6) • "It's about they're wanting to know what their individual risk is so rather than giving generic information, trying to make it as personalized as possible." (Health knowledge expert, P4)	• Module 2: builds foundational knowledge of brain anatomy and stroke types, explaining how strokes affect brain function. • Module 3: explores childhood and perinatal stroke, stroke epidemiology, highlighting its prevalence, impact, and preventability. Features health experts' videos and lived experience stories. • Module 4: introduces the FAST test and other stroke warning signs, with multilingual resources and paramedic insights. Covers emergency response, diagnosis, treatment pathways, and telestroke services, supported by survivor stories. • Module 5: covers modifiable and nonmodifiable stroke risk factors. Includes a personal risk assessment tool and minimodules on biomedical risk factors for stroke. • Module 6: covers what to expect during a stroke risk assessment appointment and includes minimodules on lifestyle-related risk factors. • Module 7: summarizes key learnings from all modules. Includes completion certificate.	MOOC content included all suggested topics, comprising 81 topics in total. One module had a readability score below 50, and 7 modules scored below 60. This poses challenges in accessibility and comprehension of these modules for people with low literacy levels.

[a]MOOC: massive open online course.

XSL•FO
RenderX

[b]FAST: face, arms, speech, and time.

Table . Participant suggestions from co-design focus groups (step 2) and comparison with MOOC[a] design (step 3) for exercise and assessment.

	Step 2: co-design focus groups			Step 3: triangulation	
Design elements	Participants' suggestion	Meaning units	Example quotes	Suggestions integrated into MOOC	Observations
Exercise	• Include risk calculator or assessments • Tailored feedback on exercises	7	• "If we want to get people to the stage of changing behavior, they have to see themselves at risk and then make that appointment to find out the degree of risk that they might be at." (Community member, P4) • "…the risk calculator is really nice and personal, and it sort of filters people down that pathway to their risk." (Health knowledge expert, P3)	Interactive activities were present in all core modules. Module 5 risk factors had cardiovascular health assessment with tailored feedback using traffic light system.	Suggestions were addressed.
Assessment	• Include assessments at the end of the modules • Avoid formal assessments	3	• "If there's activities at the end of each module that help people apply what they've learned, that may help reduce some of that frustration." (Health knowledge expert, P3)	The MOOC did not include formal assessments. End of module quiz present at the end of all modules.	Suggestions were addressed.

[a]MOOC: massive open online course.

XSL•FO
RenderX

Table . Participant suggestions from co-design focus groups (step 2) and comparison with MOOC[a] design (step 3) for media selection.

Design elements	Step 2: co-design focus groups			Step 3: triangulation	
	Participants' suggestions	Meaning units	Example quotes	Suggestions integrated into MOOC	Observations
Media selection	• Combine text, visuals, and videos to cater to different learning styles • Include animated videos for clear understanding • Videos with lived experience and health experts add credibility • Further information with links to resources for users who want to learn in detail • Multilingual options for learners from diverse linguistic backgrounds	73	• "I really liked the graphics. I think showing the different areas of the brain and labeling how it can impact differently depending on which area the stroke attacks is really powerful and concise way to show it." (Community member, P6) • "…it is good to have a person with a professional in there [the videos], but I also think it is good to have somebody with lived experience." (Health knowledge expert, P5) • "I think it needs to be available in different languages and potentially links to information in language as well." (Health knowledge expert, P2)	The MOOC included a mix of media including text, graphics, infographics, videos, transcripts, and hyperlinks to further information. Videos were short with an average length of 1 minute, 47 seconds.	Most suggestions were addressed. Multilingual options were not included.

[a]MOOC: massive open online course.

Learning Objectives

Two meaning units relating to learning objectives were identified. Participants emphasized the importance of setting clear expectations and guidelines at the beginning of the course.

> *I think setting really clear guidelines at the beginning... Committing to something that's going to run over weeks is a big task.* [Community member, P1]

In the final MOOC, learning objectives were clearly stated on the landing page of all relevant core modules except Module 7, which served as a completion module and did not require learning objectives. Four of 9 mini-modules did not include learning objectives. This inconsistency may pose challenges in engagement and clarity for learners starting those modules.

Structure

There were 11 meaning units relating to the structure of the MOOC. Participants discussed having a logical sequence for the MOOC modules and topics. They expressed a preference for delivering content in a sequence of concepts beginning with foundational concepts and building upon them, followed by optional modules that learners can tailor depending on their perceived relevance.

> *I was thinking that maybe the core content probably might be best to have that delivered in order for everybody… as people are going through them, the knowledge is layering up… Then when the individualized modules come up… they choose the module that they're really the most interested in…* [Health knowledge expert, P1]

The MOOC adopted a logical, linear structure across modules and within module topics. Core Module 5 on risk factors and Core Module 6 about action plans incorporated mini-modules that allowed learners to select topics relevant to them. The overall structure of the MOOC reflected the design discussed during the co-design sessions, supporting a logical learning structure and personalization.

Content

Content was most frequently discussed, with a total of 127 meaning units reflecting participants' suggestions relating to content. Participants identified key content topics to include in the MOOC, such as types of stroke, statistics, signs and

XSL•FO
RenderX

symptoms, and risk factors for stroke. They suggested contextualizing information to support comprehension and highlighting the relevance of the content to learners through personalization and lived experience stories. Participants also recommended presenting content in an aphasia-friendly format, using plain language, and aligning with health literacy principles.

> *So, there's a thing called universal precautions, which basically you assume everyone has low health literacy, and that way everyone can understand it; and part of that is using short, simple terms where you can.*. [Health knowledge expert, P1]

The MOOC incorporated content topics suggested by participants (Multimedia Appendix 3 for a list of content covered in the MOOC). Content across all modules was presented in simple language, supported by detailed explanations and examples provided through text and videos to foster understanding. Where technical terms were used in the content, they were bolded and detailed in a glossary section. The glossary consisted of 113 terms with simple definitions. The word count of modules ranged from 1688 to 4640 words, and readability scores ranged from 49 to 69 (equivalent to Australian Grade 8 to 12). The average readability score across modules was 60.7 (out of 100), with higher scores indicating easier readability. Seven of 16 modules scored below 60, with one mini-module (Cholesterol) scoring below 50, indicating that the content was relatively complex and may have included more technical language. Inconsistencies in readability scores may limit comprehension for learners with lower literacy. While an aphasia-friendly format or toggle was recommended, the MOOC did not include this feature.

Exercise

A total of 7 meaning units relating to exercises were identified. Participants suggested incorporating short and engaging interactive activities to enhance learner engagement.

> *Having smaller bits throughout that are gamified or fun to do…something that's less formal I think is going to be much more engaging for people, especially those with lower health literacy.* [Health knowledge expert, P3]

Six interactive activities were included across all relevant core modules, including gamified crosswords, flashcards, and a cardiovascular risk assessment with tailored feedback. These features were not present in the mini-modules.

Assessment

Overall, 3 meaning units relating to assessment were identified. Rather than including formal assessments in the MOOC, participants preferred quizzes at the end of modules to allow learners to check their understanding without added pressure.

> *I think having a more formal assessment at the end is going to be quite off-putting for people who don't pass...* [Health knowledge expert, P3]

The MOOC included 14 end-of-module quizzes across all relevant core modules and optional mini-modules, with an average of 3 questions per quiz.

Media Selection

A total of 73 meaning units relating to media selection were identified. Participants recommended using a diverse range of formats to deliver information. These included simple infographics, diagrams, short videos with minimal distractions, videos featuring both people with lived experience and health experts, and hyperlinks to additional resources.

> *With a mix of the text as well and then somebody talking, so you've got three different learning styles addressed there.* [Health knowledge expert, P2]

The MOOC incorporated a mix of media across all modules, including graphics, infographics, videos, transcripts, and hyperlinks. It included 30 videos featuring subject area experts (eg, epidemiologists, neurologists, cardiologists, paramedics, and dietitians), 28 videos featuring individuals with lived experience of stroke and carers, and 2 hand-drawn animated videos. Videos were concise, with an average duration of 1 minute and 47 seconds. While most suggestions were incorporated into the MOOC, the absence of multilingual options or features limited its accessibility to people from culturally and linguistically diverse (CALD) backgrounds. Notably, this was already identified as the next step in the broader project after evidence is established for this MOOC.

Discussion

Principal Findings

We describe the co-design of a MOOC, including an evaluation of the co-design process, aimed at improving knowledge of stroke and supporting the identification and management of stroke risk factors in the community. To the best of our knowledge, the Preventing Stroke MOOC is the first MOOC worldwide for primary prevention of stroke. The process involved active multiphase co-design with health knowledge experts, community members, and the design team. The co-design process identified topics to support a comprehensive understanding of stroke, along with delivery modes and engagement features to ensure accessibility and relevance to the intended audience. The Preventing Stroke MOOC, which included 7 core modules and 9 mini-modules, was viewed positively by participants and perceived as a useful tool for improving stroke knowledge. Through evaluating the co-design process of the final MOOC features against the original co-design recommendations, we identified gaps to inform future refinements to the MOOC.

Previous research identifies low engagement and completion as major challenges for MOOCs [36]. These challenges are often attributed to limited consideration of learners' individual needs in MOOC design, thus limiting its effectiveness [37]. Claflin et al [19] noted that engaging stakeholders early improves the quality of a MOOC and fosters success. We used co-design to develop the Preventing Stroke MOOC, involving diverse participant groups throughout the process. This ensured that the course was relevant to the needs of the intended audience and fit for purpose.

Involving participants ensured that desired features of the MOOC were identified and tailored to diverse learning needs.

XSL•FO
RenderX

Participants suggested including interactive activities and multimodal content delivery to accommodate different learning preferences. These suggestions highlight the importance of designing educational content that is accessible to learners with diverse educational needs. This aligns with existing studies on MOOC engagement, which emphasize the use of interactive features and multimodal content to support motivation, retention, and completion across literacy levels [38,39]. Our findings confirm evidence from the MOOC literature regarding learners' preferences, demonstrating that interactive activities and personalized learning are desired features of MOOCs [40,41].

Research using co-design methods in stroke has primarily focused on secondary prevention and rehabilitation, making our study a unique contribution [42]. Previous studies have mostly used co-design to develop interventions for individual cardiovascular risk factors such as physical inactivity and diet [18,43]. Our inclusion of all major risk factors for stroke in the MOOC is supported by evidence that interventions targeting multiple risk factors have a greater impact on reducing cardiovascular risk in both general and at-risk populations [44].

The evaluation of the co-design process demonstrated that most participants' suggestions from the co-design phase were reflected in the final MOOC design. Although co-design methodologies are increasingly used to develop health interventions "with" and "for" end users, the process is rarely evaluated in terms of whether the end product reflected the needs and suggestions of co-design participants [21]. Retrospective evaluation is essential to understand how the co-design process shaped the end product, why some suggestions were not addressed, and how future co-design could be improved [21]. Our evaluation of the co-design process addresses a recognized gap in the literature, as previous studies have relied on user feedback or researchers' reflections on the design process [21]. By providing a practical and transparent approach for retrospectively examining how participants' contributions are reflected in the final intervention design, this study contributes to the growing body of literature on evaluating the co-design process in health intervention development.

Evidence suggests that many MOOCs lack coherent instructional design principles, such as learning objectives and activities, which limit their effectiveness in achieving intended learning outcomes [37]. Findings from our evaluation of the co-design process informed refinements to the MOOC design prior to its public release.

Suggestions from participants that were not reflected in the MOOC included consistent learning objectives across mini-modules, an aphasia-friendly format, and multilingual options. Inconsistencies in learning objectives may affect learners' clarity about what to expect when completing these modules. Learning objectives were mostly absent from the mini-modules, which focused on specific individual risk factors and action plans, and therefore may not have required the same structured learning objectives as the core modules, which covered multiple topics. However, inconsistencies should be addressed to ensure coherence.

Co-design participants emphasized using plain language and content aligned with health literacy principles to support comprehension across different literacy levels. Readability scores varied across individual modules, with 7 modules (of 16) scoring below 60, indicating fairly difficult readability for the general population (grade 10 - 12, aged 15 to 18 years) [45]. Efforts were taken to simplify language by engaging a professional writer who adapted the draft content of the MOOC into accessible language. Through retrospective evaluation of the co-design process, conducted separately from the design and implementation process, the analysis identified a mismatch between participants' emphasis on plain language and the readability of the final MOOC content. One contributing factor may be the inclusion of medical terminology, which is often multisyllabic and difficult to further simplify without compromising accuracy. To mitigate this, the MOOC included a glossary section and features such as hover-over definitions for technical terms, thereby providing simplified definitions to support understanding of complex terminology.

Modules with low readability scores may present challenges for learners with low literacy in terms of engaging with and understanding the content [45]. This highlights important accessibility considerations for groups with lower education levels, lower socioeconomic status, and CALD populations, who are also at greater risk of stroke and may benefit most from the MOOC [46]. An important consideration in disseminating health information is to ensure that it is accessible to the broader community [47]. This is particularly relevant in primary prevention, where individuals often have limited motivation to engage with health information [48]. As the purpose of the Preventing Stroke MOOC is to educate the general community with evidence-based and accessible information, it is important to ensure that the content accommodates learners with diverse literacy levels.

Aphasia-friendly formatting was not included in the MOOC. Language deficits such as aphasia are common among stroke survivors but not among the general population [49], so this functionality was of lower priority. However, we agree that such adaptation would be necessary for digital platforms aimed at secondary stroke prevention [49]. Multilingual options within the MOOC were not incorporated, mostly due to time and cost constraints. However, it will be considered for the future to make the platform accessible to people who speak languages other than English.

Strengths and Limitations

An important strength of the study is the involvement of stakeholders in co-designing the MOOC throughout the design process. This approach aligns with person-based methods and extends beyond consultation and acceptability of the intervention [50]. Furthermore, the co-design focus groups were conducted separately for the health knowledge experts and community member groups, thereby helping to mitigate power imbalances. By adopting a comprehensive approach to reporting the co-design process across 3 phases, we provide valuable insights into planning and designing a MOOC for stroke prevention, an approach that could also be adopted to co-design MOOCs across various complex health interventions.

Our study has some limitations to consider. Our co-design participant groups did not broadly represent the Australian

XSL•FO
RenderX

population, and insights from CALD groups were not captured as this was beyond the scope of this initial development. Given that Australia is a multicultural country, our future work will focus on making the Preventing Stroke MOOC accessible to various CALD groups with limited English proficiency and literacy levels.

Community member participants in our study primarily comprised stroke survivors. While we attempted to recruit more broadly from the general community, participation from individuals without a history of stroke was limited, with only 25% (3/12) of community participants having no personal history of stroke. This reflects a broader challenge in engaging individuals at risk of stroke, who often have limited awareness or risk perception, including optimism bias, and may therefore be less likely to engage in prevention initiatives [51-53].

To ensure the relevance of the MOOC content to primary stroke prevention, co-design discussions were framed around primary prevention. Participants with lived experience of stroke contributed insights into missed opportunities for prevention, including gaps in awareness of risk factors, recognition of stroke symptoms, and early preventive action [53,54]. This was also informed by several of these people being volunteers who delivered stroke prevention talks in the community, giving them additional insights about knowledge and awareness among people without stroke. These insights informed the design of the content that aims to improve stroke knowledge, reduce false reassurance among individuals who may underestimate their risk, and encourage engagement with preventive behaviors. However, we acknowledge that these perspectives may differ from those of individuals without direct experience of stroke and should be considered when interpreting the findings. Participation was also limited to a relatively small, self-selected group, which may have included individuals who were more engaged or health-aware than the broader community. For example, community member participants without a history of stroke were carers or individuals interested in stroke prevention who regularly volunteered in community stroke prevention programs and therefore may have had greater awareness of stroke and interest in prevention. Additional strategies may be needed to better engage individuals at risk of stroke in co-design in the future.

In this study, sample adequacy was considered using the concept of information power rather than data saturation [55]. Information power emphasizes the relevance and richness of information contributed in relation to the study aim when considering sample adequacy [55]. Consistent with this approach, the focus was on obtaining rich and relevant insights throughout the co-design process.

We acknowledge the potential limitations of a single researcher conducting the coding, which meant that interrater reliability could not be assessed. While the use of design elements adapted from the ADDIE model in the final stage provided a structured and consistent approach for organizing and comparing findings to reduce variance, we cannot eliminate the possibility of variation in coding. We believe that the primary aim of interpreting and comparing participants' input with the MOOC design should be considered when interpreting the findings. We also note ongoing debate in qualitative research regarding the role of multiple coders, particularly in interpretive approaches where emphasis is placed on depth of understanding [35]. We mitigated risks associated with one coder by having iterative discussions within the research team to review and refine interpretations to enhance the trustworthiness of the results [29,33].

There is currently no established guidance on retrospectively evaluating the co-design process of an end product. Future research is warranted to establish guidelines for rigorous retrospective evaluation of the co-design process to ensure the end product reflects the needs, priorities, and lived experience of all stakeholders.

Conclusion

Through a comprehensive co-design process, we have developed a Preventing Stroke MOOC based on suggestions from key stakeholders. It offers great potential to educate the community about stroke and support the recognition and management of stroke risk factors. Overall, this study provides important insights into the educational information, design features, and delivery approaches considered important by stakeholders for primary stroke prevention. In our future work, we will assess the acceptability, engagement, and effectiveness of the Preventing Stroke MOOC in improving stroke knowledge and the uptake of healthy behaviors in a fully powered randomized controlled trial before its public release in 2026.

Acknowledgments

We acknowledge and thank all participants who contributed their time and insight into shaping the design features and content of the Preventing Stroke MOOC. We thank Jacqueline Fox for her support in refining the language and improving clarity of the content. We also thank the Digital Futures team for their assistance in developing the Preventing Stroke MOOC platform. We acknowledge people with lived experience of stroke and their support people.

Funding

Love Your Brain is a stroke prevention digital platform that is a collaborative research project between Monash University, the Menzies Institute for Medical Research, and the Stroke foundation. The Australian Government Medical Research Future Fund (MRFF #2015976) funded this project. Investigators on the MRFF grant include MKF, SLG, TP, JEB, BB, DAC, JC, EH, TK, LM, MTO and AGT. Research fellowship support from the National Heart Foundation was awarded to MFK, SLG, JEB and LLD (105737, 108523, 102061, 104751 and 107125). AG is a PhD student supported by the Tasmania Graduate Research

XSL•FO
RenderX

Scholarship (TGRS). MFK, SLG, TP, BB, DAC, RK, MRN, AGT also participate as investigators in the Synergy Grant funded by the National Health and Medical Research Council of Australia (STOPstroke #1182071). The funders had no role in the design, conduct, analysis or reporting of this study.

Data Availability

Data for this study are not publicly available because participants consented only to the inclusion of nonidentifiable excerpts in publications and did not consent to the distribution of their data.

Authors' Contributions

Conceptualization: SLG, SB, MFK
Data curation: TP, SB, CB, RFP
Formal analysis: AG
Funding acquisition: SLG, MFK, TP, DAC, BB, JEB, JC, EH, TK, LM, MRN, MTO, AGT
Investigation: SLG, MFK, TP, CB, CF, SB, RFP
Methodology: SLG, MFK, SB, TP, RFP, AG
Project administration: CF, CB, RFP
Supervision: SLG, SB, REC
Writing–original draft: AG
Writing–review & editing: AG, SLG, SB, REC, MFK, TP, CB, AGT, LLD, MRN, LM, CF, DAC, JC, SR

Conflicts of Interest

MFK and DAC are members of the Australian Stroke Clinical Registry Management Committee. AGT is a former Board Member of the Stroke Foundation. MFK is a Deputy Chair of the Stroke Foundation Research Advisory Committee. SLG chairs the Stroke Prevention Advisory Committee and is a member of its Clinical Council. MRN is a member of a Novartis lipids advisory board, unrelated to the submitted work.

Multimedia Appendix 1
Pilot module user testing focus group discussion guide.
[DOCX File, 15 KB - cardio_v10i1e89358_app1.docx]

Multimedia Appendix 2
Co-design phase focus group discussion topics relevant to the MOOC design.
[DOCX File, 18 KB - cardio_v10i1e89358_app2.docx]

Multimedia Appendix 3
Preventing Stroke MOOC design features.
[DOCX File, 19 KB - cardio_v10i1e89358_app3.docx]

Multimedia Appendix 4
Example of analytic process.
[PDF File, 60 KB - cardio_v10i1e89358_app4.pdf]

Checklist 1
COREQ checklist.
[DOCX File, 25 KB - cardio_v10i1e89358_app5.docx]

References

1. O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet 2016 Aug 20;388(10046):761-775. [doi: 10.1016/S0140-6736(16)30506-2] [Medline: 27431356]
2. Choudhury MJH, Chowdhury MTI, Nayeem A, Jahan WA. Modifiable and non-modifiable risk factors of stroke: a review update. J Natl Inst Neurosci Bangladesh 2015;1(1):22-26. [doi: 10.3329/jninb.v1i1.22944]
3. Heart, stroke and vascular disease: Australian facts. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts/contents/summary [accessed 2026-04-05]
4. Pandian JD, Gall SL, Kate MP, et al. Prevention of stroke: a global perspective. Lancet 2018 Oct 6;392(10154):1269-1278. [doi: 10.1016/S0140-6736(18)31269-8] [Medline: 30319114]

XSL•FO
RenderX

5. Hartigan I, O'Connell E, O'Brien S, et al. The Irish national stroke awareness campaign: a stroke of success? Appl Nurs Res 2014 Nov;27(4):e13-e19. [doi: 10.1016/j.apnr.2014.05.004] [Medline: 25015898]
6. Lecouturier J, Rodgers H, Murtagh MJ, White M, Ford GA, Thomson RG. Systematic review of mass media interventions designed to improve public recognition of stroke symptoms, emergency response and early treatment. BMC Public Health 2010 Dec 23;10:784. [doi: 10.1186/1471-2458-10-784] [Medline: 21182777]
7. Kilkenny MF, Purvis T, Werner M, et al. Improving stroke knowledge through a "volunteer-led" community education program in Australia. Prev Med 2016 May;86:1-5. [doi: 10.1016/j.ypmed.2016.01.015] [Medline: 26820114]
8. Bray JE, Johnson R, Trobbiani K, et al. Australian public's awareness of stroke warning signs improves after national multimedia campaigns. Stroke 2013 Dec;44(12):3540-3543. [doi: 10.1161/STROKEAHA.113.002987] [Medline: 24135926]
9. Bray JE, Mosley I, Bailey M, Barger B, Bladin C. Stroke public awareness campaigns have increased ambulance dispatches for stroke in Melbourne, Australia. Stroke 2011 Aug;42(8):2154-2157. [doi: 10.1161/STROKEAHA.110.612036] [Medline: 21757668]
10. Australians are not FAST enough. Stroke Foundation. URL: https://strokefoundation.org.au/media-centre/media-releases/2021/11/australians-are-not-fast-enough [accessed 2025-04-25]
11. Kohler S. Can internet access growth help reduce the global burden of noncommunicable diseases? Online J Public Health Inform 2013;5(2):221. [doi: 10.5210/ojphi.v5i2.4667] [Medline: 23923103]
12. Pilli O, Admiraal W. A taxonomy of massive open online courses. Contemp Educ Technol 2016;7(3):223-240. [doi: 10.30935/cedtech/6174]
13. Baturay MH. An overview of the world of MOOCs. Procedia Soc Behav Sci 2015 Feb;174:427-433. [doi: 10.1016/j.sbspro.2015.01.685]
14. Martinez Escobedo I, Doherty K, Eccleston C. Health promotion MOOCs (hpMOOCs): a dual lens for assessing quality. Am J Distance Educ 2024 Apr 2;38(2):168-184. [doi: 10.1080/08923647.2024.2325845]
15. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 7;22(1):877. [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
16. Steen M. Co-design as a process of joint inquiry and imagination. Des Issues 2013 Apr;29(2):16-28. [doi: 10.1162/DESI_a_00207]
17. Talevski J, Kulnik ST, Jessup RL, Falls R, Cvetanovska N, Beauchamp A. Use of co-design methodology in the development of cardiovascular disease secondary prevention interventions: a scoping review. Health Expect 2023 Feb;26(1):16-29. [doi: 10.1111/hex.13633] [Medline: 36366855]
18. McGill B, Corbett L, Grunseit AC, Irving M, O'Hara BJ. Co-produce, co-design, co-create, or co-construct—who does it and how is it done in chronic disease prevention? A scoping review. Healthcare (Basel) 2022;10(4):647. [doi: 10.3390/healthcare10040647]
19. Claflin SB, Gates R, Maher M, Taylor BV. Building a successful massive open online course about multiple sclerosis: a process description. J Med Internet Res 2020 Jul 29;22(7):e16687. [doi: 10.2196/16687] [Medline: 32723726]
20. Dratsiou I, Varella A, Gkrimoura A, et al. Lessons learned co-creating a Massive Open Online Course (MOOC) on refugees' health integration. Int J Changes Educ 2025;3(3):255-267. [doi: 10.47852/bonviewIJCE52024606]
21. Peters S, Guccione L, Francis J, et al. Evaluation of research co-design in health: a systematic overview of reviews and development of a framework. Implement Sci 2024 Sep 11;19(1):63. [doi: 10.1186/s13012-024-01394-4] [Medline: 39261956]
22. Kilkenny MF, Cadilhac DA, Thrift AG, et al. Evaluation of a digital health platform for preventing stroke in the Australian community: study protocol for a randomized controlled trial - Love Your Brain. PLoS One 2025;20(9):e0330868. [doi: 10.1371/journal.pone.0330868] [Medline: 40906720]
23. Burns C, Kilkenny MF, Purvis T, et al. Processes underpinning successful co-design: lessons from a digital health project. Health Expect 2025 Jun;28(3):e70272. [doi: 10.1111/hex.70272] [Medline: 40405447]
24. Purvis T, Burns C, Barker S, et al. Co-designing a digital stroke prevention platform: leveraging lived experience and expert advice. Health Expect 2025 Jun;28(3):e70293. [doi: 10.1111/hex.70293] [Medline: 40405455]
25. Kilkenny MF, Freak-Poli R, Burns C, et al. Development of a health text message system to support stroke prevention: a component of the Love Your Brain digital platform. Health Expect 2025 Oct;28(5):e70471. [doi: 10.1111/hex.70471] [Medline: 41117759]
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Bird M, McGillion M, Chambers EM, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 1;7(1):12. [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
28. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
29. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]

XSL•FO
RenderX

30. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
31. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
32. Kurt S. ADDIE Model: instructional design. Educational Technology. 2018 Dec 16. URL: https://educationaltechnology.net/the-addie-model-instructional-design/ [accessed 2025-04-22]
33. Kyngäs H, Kääriäinen M, Elo S. The trustworthiness of content analysis. In: Kyngäs H, Mikkonen K, Kääriäinen M, editors. The Application of Content Analysis in Nursing Science Research: Springer; 2020:41-48. [doi: 10.1007/978-3-030-30199-6_5]
34. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research: norms and guidelines for CSCW and HCI practice. Proc ACM Hum-Comput Interact 2019;3(CSCW):1-23. [doi: 10.1145/3359174]
35. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 1;19. [doi: 10.1177/1609406919899220]
36. Jordan K. Initial trends in enrolment and completion of massive open online courses. Int Rev Res Open Distrib Learn 2014;15(1):133-160. [doi: 10.19173/irrodl.v15i1.1651]
37. Pilli O, Admiraal W. Students' learning outcomes in Massive Open Online Courses (MOOCs): some suggestions for course design. Yuksekogretim Derg 2017 Apr 1;7(1):46-71. [doi: 10.2399/yod.17.001]
38. Rodrigues Major R, Mira da Silva M. Evaluating a gamified MOOC. Cogent Educ 2025 Dec 31;12(1). [doi: 10.1080/2331186X.2025.2479400]
39. Iniesto F, Rodrigo C. Understanding accessibility in MOOCs: findings and recommendations for future designs. J Interact Media Educ 2024 Sep 20;2024(1):20. [doi: 10.5334/jime.897]
40. Oh EG, Cho MH, Chang Y. Learners' perspectives on MOOC design. Distance Educ 2023 Jul 3;44(3):476-494. [doi: 10.1080/01587919.2022.2150126]
41. Nanda G, A. Douglas K, R. Waller D, E. Merzdorf H, Goldwasser D. Analyzing large collections of open-ended feedback from MOOC learners using LDA topic modeling and qualitative analysis. IEEE Trans Learn Technol 2021;14(2):146-160. [doi: 10.1109/TLT.2021.3064798]
42. Singh RJ, Chen S, Ganesh A, Hill MD. Long-term neurological, vascular, and mortality outcomes after stroke. Int J Stroke 2018 Oct;13(8):787-796. [doi: 10.1177/1747493018798526] [Medline: 30160619]
43. Tay BSJ, Edney SM, Brinkworth GD, et al. Co-design of a digital dietary intervention for adults at risk of type 2 diabetes. BMC Public Health 2021 Nov 11;21(1). [doi: 10.1186/s12889-021-12102-y] [Medline: 34763701]
44. Sisti LG, Dajko M, Campanella P, Shkurti E, Ricciardi W, de Waure C. The effect of multifactorial lifestyle interventions on cardiovascular risk factors: a systematic review and meta-analysis of trials conducted in the general population and high risk groups. Prev Med 2018 Apr;109:82-97. [doi: 10.1016/j.ypmed.2017.12.027] [Medline: 29291422]
45. Cheng C, Dunn M. Health literacy and the Internet: a study on the readability of Australian online health information. Aust N Z J Public Health 2015 Aug;39(4):309-314. [doi: 10.1111/1753-6405.12341] [Medline: 25716142]
46. Avan A, Digaleh H, Di Napoli M, et al. Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: an ecological analysis from the Global Burden of Disease Study 2017. BMC Med 2019 Oct 24;17(1). [doi: 10.1186/s12916-019-1397-3] [Medline: 31647003]
47. Schillinger D. Social determinants, health literacy, and disparities: intersections and controversies. Health Lit Res Pract 2021 Jul;5(3):e234-e243. [doi: 10.3928/24748307-20210712-01] [Medline: 34379549]
48. Gonçalves-Fernández ML, Pino-Juste M. Health literacy in healthy adults: a systematic review of recent evidence. Aten Primaria 2025 Nov;57(11):103300. [doi: 10.1016/j.aprim.2025.103300] [Medline: 40680630]
49. Rose T, Worrall L, McKenna K. The effectiveness of aphasia‐friendly principles for printed health education materials for people with aphasia following stroke. Aphasiology 2003 Jan;17(10):947-963. [doi: 10.1080/02687030344000319]
50. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
51. Sullivan KA, White KM, Young RM, Scott C, Mulgrew K. Developing a stroke intervention program: what do people at risk of stroke want? Patient Educ Couns 2008 Jan;70(1):126-134. [doi: 10.1016/j.pec.2007.09.006] [Medline: 17997262]
52. Kitko L, Hupcey JE. Factors that influence health-seeking behaviors of patients experiencing acute stroke. J Neurosci Nurs 2008 Dec;40(6):333-340. [doi: 10.1097/01376517-200812000-00004] [Medline: 19170299]
53. Yoon SS, Byles J. Perceptions of stroke in the general public and patients with stroke: a qualitative study. BMJ 2002 May 4;324(7345):1065-1068. [doi: 10.1136/bmj.324.7345.1065] [Medline: 11991910]
54. Nelson PA, Kane K, Chisholm A, et al. "I should have taken that further" - missed opportunities during cardiovascular risk assessment in patients with psoriasis in UK primary care settings: a mixed-methods study. Health Expect 2016 Oct;19(5):1121-1137. [doi: 10.1111/hex.12404] [Medline: 26340682]
55. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]

XSL•FO
RenderX

Abbreviations

ADDIE: Analysis, Design, Development, Implementation, and Evaluation
CALD: culturally and linguistically diverse
COREQ: Consolidated Criteria for Reporting Qualitative Research
CRG: community reference group
FAST: face, arms, speech, and time
MCQ: multiple-choice question
MIMR: Menzies Institute for Medical Research
MOOC: massive open online course
UTAS: University of Tasmania

Edited by J Rivers; submitted 15.Dec.2025; peer-reviewed by KA Baxter, LN Mayigane; revised version received 31.Aug.2026; accepted 31.Aug.2026; published 20.Oct.2026.

Please cite as:
Gurung A, Barker S, Rehman S, Climie RE, Freak-Poli R, Burns C, Kilkenny MF, Purvis T, Farmer C, Booth B, Bray JE, Cadilhac DA, Cameron J, Dalli LL, Ho S, Horton E, Kleinig T, Murphy L, Nelson MR, Olaiya MT, Thrift AG, Gall SL
A Massive Open Online Course for Primary Prevention of Stroke: Evaluation of Co-Design Process
JMIR Cardio 2026;10:e89358
URL: https://cardio.jmir.org/2026/1/e89358
doi:10.2196/89358

© Aastha Gurung, Seamus Barker, Sabah Rehman, Rachel E Climie, Rosanne Freak-Poli, Catherine Burns, Monique F Kilkenny, Tara Purvis, Christine Farmer, Brenda Booth, Janet E Bray, Dominique A Cadilhac, Jan Cameron, Lachlan L Dalli, Stephanie Ho, Eleanor Horton, Timothy Kleinig, Lisa Murphy, Mark R Nelson, Muideen T Olaiya, Amanda G Thrift, Seana L Gall. Originally published in JMIR Cardio (https://cardio.jmir.org), 20.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Machine Learning Models for Mortality Prediction in Intensive Care Unit Patients With Ischemic Stroke Associated With Intracranial Artery Stenosis: Retrospective Cohort Study

Kun Zhang[1], MD; Ruomeng Chen[1], MD; Jingyi Yang[2], MSc; Yan Yan[1], MD; Lijuan Liu[1], MD; Chaoyue Meng[1], MD; Peifang Li[3], MD; Guoying Xing[4], MD; Xiaoyun Liu[1], MD

[1]The First Hospital of Hebei Medical University, No. 89, Donggang Rd, Yuhua District, Shijiazhuang, Hebei, China

[2]The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

[3]Handan Central Hospital, Handan, Hebei, China

[4]Zhaoxian Traditional Chinese Medicine Hospital, Shijiazhuang, Hebei, China

Corresponding Author:
Xiaoyun Liu, MD
The First Hospital of Hebei Medical University, No. 89, Donggang Rd, Yuhua District, Shijiazhuang, Hebei, China

Abstract

Background: Mortality prediction in intensive care unit (ICU) patients with ischemic stroke complicated by intracranial artery stenosis or occlusion remains difficult. Conventional scoring systems often lack discriminatory power and fail to provide individualized risk estimates. Machine learning approaches have been increasingly explored to integrate diverse clinical features for prognostic modeling.

Objective: This study aims to develop and evaluate machine learning models for individualized mortality prediction in ICU patients with ischemic stroke associated with intracranial artery stenosis or occlusion.

Methods: Using the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, we conducted a retrospective cohort study including 5280 adult ICU patients identified through *International Classification of Diseases, Ninth and Tenth Revision* (*ICD-9/10*) codes. Mortality status was determined based on the presence of a recorded date of death (dod) in the MIMIC-IV database. Patients with a documented dod were classified as deceased, whereas those without a recorded dod were classified as nondeceased. The primary outcome was all-cause mortality as recorded in the MIMIC-IV database, defined by the presence of a documented dod. Patients were randomly split into training (n=3696, 70%) and testing (n=1584, 30%) cohorts. Missing value imputation, correlation reduction, and multistep supervised feature selection (gradient boosting, BorutaShap, recursive feature elimination with cross-validation, LassoCV, and chi-square analysis) were performed exclusively within the training set and subsequently applied to the test set, resulting in 35 retained predictive features. Eight machine learning models—including light gradient boosting machine (LightGBM), Bagging (bootstrap aggregating), random forest, logistic regression, support vector machine, gradient boosting, adaptive boosting, and k-nearest neighbors—were trained with hyperparameter optimization using RandomizedSearchCV. Model performance was evaluated using area under the curve, accuracy, recall, precision, F_1-score, and calibration curves. Shapley additive explanations were used for global and individual-level interpretability.

Results: LightGBM, Bagging, and logistic regression demonstrated comparable discrimination, achieving an area under the curve of approximately 0.82 - 0.83 and accuracy above 73% on the independent test set. LightGBM demonstrated balanced performance (recall 0.70; precision 0.72) and good calibration. Shapley additive explanations analysis identified acute physiology score III, suspected infection, Charlson comorbidity index, age, weight on admission, and red cell distribution width as the most influential predictors. Overall, higher physiological severity, greater comorbidity burden, and older age were consistently associated with increased observed mortality risk.

Conclusions: Machine learning models—including LightGBM and Bagging—provide interpretable predictions of all-cause mortality in ICU patients with ischemic stroke and intracranial arterial disease. These models highlight key prognostic features and may support mortality risk stratification. External validation and evaluation of workflow integration are warranted before clinical implementation.

(JMIR Cardio 2026;10:e82042) doi:10.2196/82042

KEYWORDS

ischemic stroke; intracranial arterial stenosis; critical care; machine learning; mortality prediction; explainable artificial intelligence; intensive care unit; ICU outcomes

XSL•FO
RenderX

Introduction

Cerebrovascular diseases, particularly those associated with intracranial artery stenosis or occlusion, are major contributors to morbidity and mortality among critically ill patients [1]. In the intensive care unit (ICU), these conditions often lead to acute neurological deterioration, secondary complications, and poor clinical outcomes, posing significant management challenges. Despite the widespread use of traditional clinical scoring systems, accurate and individualized mortality prediction remains difficult.

Recent advances in machine learning (ML) offer promising approaches to address this challenge [2]. ML algorithms can process large-scale, high-dimensional clinical data to detect complex, nonlinear relationships that conventional methods may overlook [3]. Such tools have shown potential in critical care settings, particularly for mortality prediction, risk stratification, and outcome modeling in severe illnesses including stroke and sepsis.

The Medical Information Mart for Intensive Care IV (MIMIC-IV) database [4] is a publicly accessible ICU dataset containing detailed records of over 350,000 ICU admissions, including demographics, diagnoses, physiological measurements, laboratory tests, and treatments. This rich resource facilitates the development of robust, data-driven predictive models [5].

In this study, we leveraged the MIMIC-IV database to develop and evaluate ML models for predicting in-hospital and post-discharge mortality in ICU patients diagnosed with ischemic stroke associated with intracranial artery stenosis or occlusion. We used advanced feature selection techniques—including gradient boosting classifier, BorutaShap, recursive feature elimination with cross-validation (RFECV), and LassoCV—to optimize the input variables [6]. Additionally, Shapley additive explanations (SHAPs) were integrated to enhance model interpretability and transparency, providing insight into the contribution of each feature to individual predictions [7].

Although several ML-based models have been developed for predicting outcomes in patients with ischemic stroke [8-10], most existing studies have focused on general stroke cohorts or emergency department populations and have rarely examined critically ill patients with intracranial artery stenosis or occlusion. Moreover, prior models typically target in-hospital mortality alone, rely on limited variable sets, or lack transparent feature-interpretation methods. In contrast, our study incorporates a large and granular ICU cohort from the MIMIC-IV database, applies a multistep feature selection pipeline, evaluates multiple ML algorithms under a unified framework, and uses SHAP to provide clinically interpretable insights. By including both in-hospital and postdischarge deaths recorded in the database, our model provides a pragmatic assessment of observed all-cause mortality risk in ICU patients, reflecting real-world outcome documentation in large critical care registries.

Our aim was to establish an interpretable and high-performing prognostic model to support timely and informed clinical decision-making for ICU patients with ischemic stroke. The following sections describe our methodology, present the results, and discuss the clinical implications and future research directions.

Methods

Study Design and Data Source

This retrospective cohort study was conducted using the publicly available MIMIC-IV database, which contains detailed clinical data from over 350,000 ICU admissions at the Beth Israel Deaconess Medical Center between 2008 and 2019. The database includes a comprehensive range of variables, such as demographic characteristics, vital signs, laboratory results, diagnoses, medications, and procedures.

Mortality outcomes were ascertained using the publicly available MIMIC-IV database (version 3.1; released on October 11, 2024). Mortality status was determined based on the presence of a recorded date of death (dod) in the database. Patients with a documented dod were classified as deceased, whereas patients without a recorded dod were classified as alive. This definition captures all-cause in-hospital mortality as well as postdischarge mortality when documented in MIMIC-IV. The primary outcome was all-cause mortality, defined by the presence of a recorded dod in the MIMIC-IV database. Patients with a documented dod were classified as deceased, whereas those without a recorded dod were classified as nondeceased. No additional time-window restriction was applied.

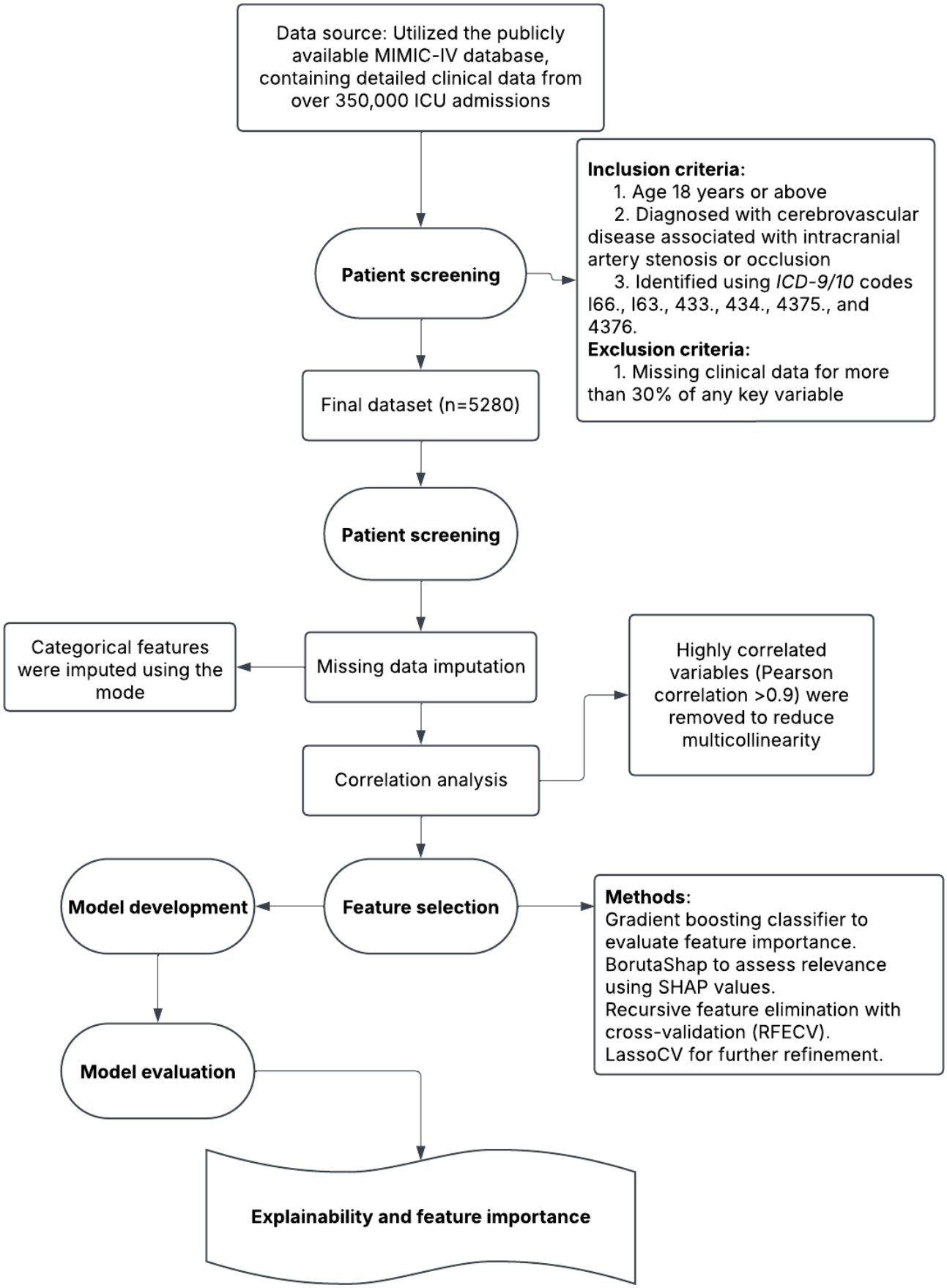
This study focused on adult ICU patients diagnosed with ischemic stroke associated with intracranial artery stenosis or occlusion. Patients were identified using *International Classification of Diseases, Ninth and Tenth Revision* (*ICD-9/10*) codes corresponding to cerebrovascular pathology, specifically: I66., I630-I637, 433., 434.*, 437.5, and 437.6. Because cohort selection was not restricted to the principal diagnosis, the study population included critically ill patients in whom ischemic stroke may have been either the primary diagnosis or a significant comorbid condition.

A total of 8437 patient records were initially identified with relevant cerebrovascular diagnoses. To ensure data integrity and minimize redundancy, only the first ICU admission for each patient was considered. After excluding individuals with multiple admissions or incomplete records, the final study cohort comprised 5280 patients aged 18 years or older.

The primary outcome was all-cause mortality, defined by the presence of a recorded dod in the MIMIC-IV database. The patient selection process is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart of the study design and methodology. *ICD-9/10*: *International Classification of Diseases, Ninth and Tenth Revision*; ICU: intensive care unit; MIMIC-IV: Medical Information Mart for Intensive Care IV; SHAP: Shapley additive explanations.



XSL•FO
RenderX

Inclusion and Exclusion Criteria

Patients were included if they met the following criteria: (1) aged 18 years or older; (2) diagnosed with ischemic stroke associated with intracranial artery stenosis or occlusion, identified using *ICD-9/10* codes from MIMIC-IV, including I66., I630-I637, 433., 434., 437.5, and 437.6; and (3) had documented mortality status based on the presence or absence of a recorded dod in the MIMIC-IV database. From all 364,627 patients in MIMIC-IV, a total of 8437 hospitalization records fulfilled the cerebrovascular diagnostic criteria. For patients with multiple admissions, only the first ICU admission was retained to avoid duplication, resulting in a final cohort of 5280 adult ICU patients.

Outcome Definition

The primary outcome was observed all-cause mortality, defined by the presence of a recorded dod in the MIMIC-IV database. This outcome reflects mortality events documented during hospitalization and after discharge when available. The absence of a recorded dod was interpreted as survival within the observation window. This definition was applied consistently across the entire cohort and all modeling stages.

Variables and Features

A comprehensive set of variables was included, comprising:

- Demographic variables: age, sex, height, and weight.
- Clinical parameters: vital signs (eg, heart rate, blood pressure, respiratory rate), laboratory findings (eg, hemoglobin, white blood cell count, glucose, calcium, creatinine).
- Scoring systems: sequential organ failure assessment, acute physiology and chronic health evaluation III (APS-III), and Glasgow Coma Scale.
- Clinical history: Clinical history variables included the Charlson comorbidity index, presence of malignancy, suspected infection (defined algorithmically based on antibiotic administration and body fluid culture collection), use of mechanical ventilation, and pharmacological treatments (eg, anticoagulants and antihypertensive agents).

To ensure adequate data quality, any variable with more than 30% missing values was excluded from further analysis. This criterion was applied to variables only, and no patients were excluded based on this step. Following preprocessing, 151 features were retained for further analysis. To ensure early risk stratification and avoid look-ahead bias, all predictive features were extracted exclusively from the early phase of ICU admission. Demographic characteristics were obtained at baseline, and all laboratory variables were defined as the first available measurement after ICU admission. Physiological parameters, clinical scores, and treatment-related indicators (including mechanical ventilation status and suspected infection) were derived from data recorded within the first 24 hours following ICU admission. No information collected beyond this predefined early time window was used for feature construction, model training, or prediction.

Data Preprocessing and Feature Selection

To minimize the risk of data leakage, all preprocessing and feature selection procedures were conducted strictly within the training data only.

The dataset was first randomly split into a training set (n=3696, 70%) and an independent testing set (n=1584, 30%). Missing data imputation was performed by fitting the imputation strategy exclusively on the training set. Continuous variables were imputed using the median values derived from the training data to reduce the influence of outliers, while categorical variables were imputed using the mode. The fitted imputation parameters were then applied unchanged to the testing set.

To reduce multicollinearity, Pearson correlation analysis was conducted within the training set. Features with a correlation coefficient greater than 0.9 were excluded, including variables such as “dbp,” “hematocrit,” “rbc,” and “mechvent_score.” The same feature exclusion rules were subsequently applied to the testing set.

Feature selection was performed using a multistep strategy applied exclusively to the training data. This process included (1) gradient boosting classifier–based feature importance ranking; (2) BorutaShap, which integrates SHAP values with the Boruta algorithm for robust supervised feature selection; (3) RFECV; (4) LassoCV for regularization-based feature shrinkage; and (5) chi-square tests to assess statistical associations (P<.05).

To enhance robustness and reduce method-specific bias, a consensus feature selection strategy was adopted. Only features consistently selected across multiple methods were retained. This approach resulted in the identification of 35 predictive features with the highest and most stable importance, including “inr_max,” “calcium,” “spo2,” “oasis_x,” “suspected_infection,” and “anchor_age.” The final selected feature set was fixed after training and applied unchanged to the independent testing set.

Validation Strategy

In addition to the conventional random split, we performed a temporal validation to assess model generalizability across different calendar periods and to minimize potential information leakage due to long study duration. Specifically, patients were grouped according to the MIMIC-IV anchor_year_group variable. Admissions from earlier calendar periods were used for model training, while more recent admissions were reserved for testing. This temporal split mimics a real-world deployment scenario, in which models trained on historical data are applied to future patients. All preprocessing steps, feature selection procedures, and hyperparameter tuning were performed exclusively within the temporally defined training set, and the finalized models were evaluated on the temporally independent test set using the same performance metrics as in the random split analysis.

Model Development and Hyperparameter Optimization

The training set was used exclusively for model development and hyperparameter tuning, while the testing set was reserved for final performance evaluation. Prior to model training, feature

XSL•FO
RenderX

standardization was performed using *z*-score normalization, with scaling parameters estimated from the training data and subsequently applied to the testing data.

Multiple ML classifiers were trained, including light gradient boosting machine (LightGBM), Bagging (bootstrap aggregating), k-nearest neighbors, logistic regression, support vector machine (SVM), random forest, adaptive boosting (AdaBoost), and gradient boosting.

Hyperparameter optimization for each model was conducted using RandomizedSearchCV within the training set, using cross-validation to identify optimal parameter combinations. This training-only optimization strategy ensured that no information from the testing data influenced model selection or tuning. All cross-validation procedures were conducted exclusively within the training set.

Model performance was evaluated using several metrics, including:

- F_1-score: The harmonic mean of precision and recall.
- Precision: The proportion of predicted positives that were positive.
- Recall: The proportion of actual positives that were correctly identified.
- Area under the receiver operating characteristic curve (ROC AUC): A performance measurement for classification problems at various threshold settings.
- Accuracy: The proportion of correctly classified instances.

The performance of each model was visualized through ROC curves to assess classification ability. Confusion matrices were also generated to assess the number of true positives, false positives, true negatives, and false negatives for each model. Calibration performance was assessed using both calibration curves and quantitative metrics, including the Brier score, calibration slope, and calibration intercept, to provide a comprehensive evaluation of the reliability of predicted probabilities.

Decision Curve Analysis

To evaluate the clinical utility and net benefit of the predictive models, we performed decision curve analysis (DCA). DCA quantifies the net benefit of using a model to guide clinical decisions across a range of threshold probabilities, which reflects the trade-off that clinicians and patients are willing to make between the benefits of a true positive and the harms of a false positive. The net benefit was calculated for our top-performing models (LightGBM, Bagging, extreme gradient boosting (XGBoost), random forest, and AdaBoost) and compared against 2 default strategies: “treat all” (assuming all patients are classified as high risk) and “treat none” (assuming no patients will die). A model with a higher net benefit across a range of thresholds is considered to have greater clinical utility.

Model Interpretability

To interpret model predictions, we applied SHAP, which quantifies the contribution of each feature to individual predictions. SHAP summary and scatter plots were generated for the best-performing model, LightGBM, offering visual insights into key mortality predictors.

Statistical Analysis

Chi-square tests were used to determine associations between categorical features and mortality, with a significance level of $P<.05$. The Kolmogorov-Smirnov test was applied to compare the distribution of continuous features between survival and mortality groups.

Ethical Considerations

This study used the data from the publicly available and fully deidentified MIMIC-IV database. Because the dataset contains no identifiable private information and all patient data are anonymized, this retrospective analysis is exempt from institutional review board approval. The use of MIMIC-IV data complies with all relevant ethical guidelines and the data use agreement required for access to the database. No direct patient contact or intervention occurred, and therefore informed consent was not required.

Results

Patient Characteristics

A total of 5280 ICU patients diagnosed with cerebrovascular disease and intracranial artery stenosis or occlusion were included. The cohort had a mean age of 69.1 years (SD 14.6), and 52.4% (n=2765) were male. Table 1 summarizes key demographic, clinical, and laboratory characteristics.

XSL•FO
RenderX

Table . Demographic, clinical, and laboratory characteristics stratified by mortality status (recorded death date vs no recorded death date)[a].

	Total (N=5280)	No recorded date of death (n=2952)	Recorded date of death (n=2328)	*P* value
Demographic characteristics				
Age (y), mean (SD)	69.09 (14.61)	66.09 (14.92)	72.89 (13.26)	<.001
Male, n (%)	2765 (52.4)	1178 (50.6)	1587 (53.8)	.02
Female, n (%)	2515 (47.6)	1150 (49.4)	1365 (46.2)	.02
Race or ethnicity, n (%)				
White	3495 (66.2)	1973 (66.8)	1522 (65.4)	—[b]
Black	533 (10.1)	264 (8.9)	269 (11.6)	—
Hispanic or Latino	156 (3)	95 (3.2)	61 (2.6)	—
Asian	144 (2.7)	82 (2.8)	62 (2.7)	—
Other or unknown	952 (18)	538 (18.2)	414 (17.8)	—
Comorbidities				
Hypertension, n (%)	1655 (31.3)	959 (32.5)	696 (29.9)	.05
Diabetes mellitus, n (%)	1799 (34)	874 (29.6)	925 (39.7)	<.001
Dyslipidemia, n (%)	1964 (37.2)	797 (34.2)	1167 (39.5)	<.001
Suspected infection, n (%)	3874 (73.4)	1876 (63.6)	1998 (85.8)	<.001
Dementia, n (%)	251 (4.8)	79 (2.7)	172 (7.4)	<.001
Malignant cancer, n (%)	438 (8.3)	118 (4)	365 (15.7)	<.001
Charlson comorbidity index, median (IQR)	6 (4-8)	5 (3-7)	7 (5-9)	<.001
Laboratory parameters				
LDL-C[c] (mg/dL), mean (SD)	85.8 (33.2)	88.6 (34.2)	82.3 (31.5)	<.001
Triglycerides (mmol/L), mean (SD)	131.3 (128.8)	130.0 (117.7)	133.0 (141.5)	.12
Blood glucose (mg/dL), mean (SD)	134.1 (60.2)	127.6 (53.0)	142.3 (67.5)	<.001
RDW[d]	14.8 (2.1)	14.4 (1.9)	15.4 (2.3)	<.001
BUN[e]	24.8 (18.5)	20.5 (13.6)	30.3 (22.1)	<.001
Clinical scores				
SOFA[f] score, mean (SD)	1.5 (1.9)	1.3 (1.8)	1.8 (2.1)	<.001
APS-III[g] score, mean (SD)	42.7 (19.9)	36.6 (16.1)	50.4 (21.4)	<.001
SAPS-II[h], mean (SD)	36.4 (13.2)	32.1 (11.3)	41.8 (13.4)	<.001
GCS[i], mean (SD)	13.8 (2.2)	14.2 (2.0)	13.7 (2.5)	<.001
In-hospital mortality				
Mortality, n (%)	669 (12.7)	0 (0)	669 (28.7)	<.001

[a]Race or ethnicity was reported as recorded in the electronic health record and aggregated for descriptive purposes only.

[b]Not available.

[c]LDL-C: low-density lipoprotein cholesterol.

[d]RDW: red cell distribution width.

[e]BUN: blood urea nitrogen.

[f]SOFA: sequential organ failure assessment.

XSL•FO
RenderX

[g]APS-III: acute physiology and chronic health evaluation III.

[h]SAPS II: Simplified Acute Physiology Score II.

[i]GCS: Glasgow Coma Scale.

The most common comorbidities were hypertension (n=1655, 31.3%), diabetes mellitus (n=1799, 34%), and dyslipidemia (n=1964, 37.2%). Based on the MIMIC-IV recorded dod, 2328 (44.1%) patients were classified as deceased, while 2952 (55.9%) patients had no recorded dod and were classified as nondeceased. The in-hospital mortality rate was 12.7% (n=669), representing a subset of deaths captured by the recorded dod in the MIMIC-IV database.

Compared with patients classified as nondeceased, those classified as deceased were significantly older (mean age: 72.9, SD 13.3 y vs mean age: 66.1, SD 14.9 y; P<.001) and had higher Charlson comorbidity index scores (median 7, IQR 5-9 vs median 5, IQR 3-7; P<.001). They also had higher incidences of suspected infection (1998/2328, 85.8% vs 1876/2952, 63.6%; P<.001) and dementia (172/2328, 7.4% vs 79/2952, 2.7%; P<.001). Group differences were observed for laboratory variables such as low-density lipoprotein cholesterol and red cell distribution width (both P<.001), whereas triglycerides did not differ significantly (P=.12).

Feature Selection

Following imputation and correlation filtering, 107 features were included. After applying the feature selection pipeline (gradient boosting, BorutaShap, RFECV, LassoCV, and Chi-square testing), 35 features were consistently identified as significant. These features were retained for model development. The full list of selected features is available in Multimedia Appendix 1.

Model Performance

Eleven ML classifiers were trained and evaluated using the selected features. Table 2 presents the performance metrics, and Figure 2 illustrates ROC curves for each model.

Table . Performance metrics of machine learning models.

Classifier	Accuracy	ROC AUC[a]	Recall	Precision	F_1-score
Logistic regression	74.18	0.82	0.65	0.73	0.69
SVM[b]	74.31	0.82	0.66	0.73	0.69
KNN[c]	72.29	0.82	0.51	0.78	0.62
Decision tree	69.70	0.75	0.67	0.65	0.66
Random forest	72.92	0.82	0.66	0.70	0.68
AdaBoost	74.12	0.82	0.67	0.72	0.69
Gradient boosting	73.30	0.81	0.65	0.72	0.68
XGBoost[d]	73.42	0.82	0.68	0.70	0.69
LightGBM[e]	74.87	0.82	0.70	0.72	0.71
Bagging	73.93	0.82	0.72	0.70	0.71
Voting	72.98	0.83	0.61	0.73	0.66

[a]ROC AUC: area under the receiver operating characteristic curve.

[b]SVM: support vector machine.

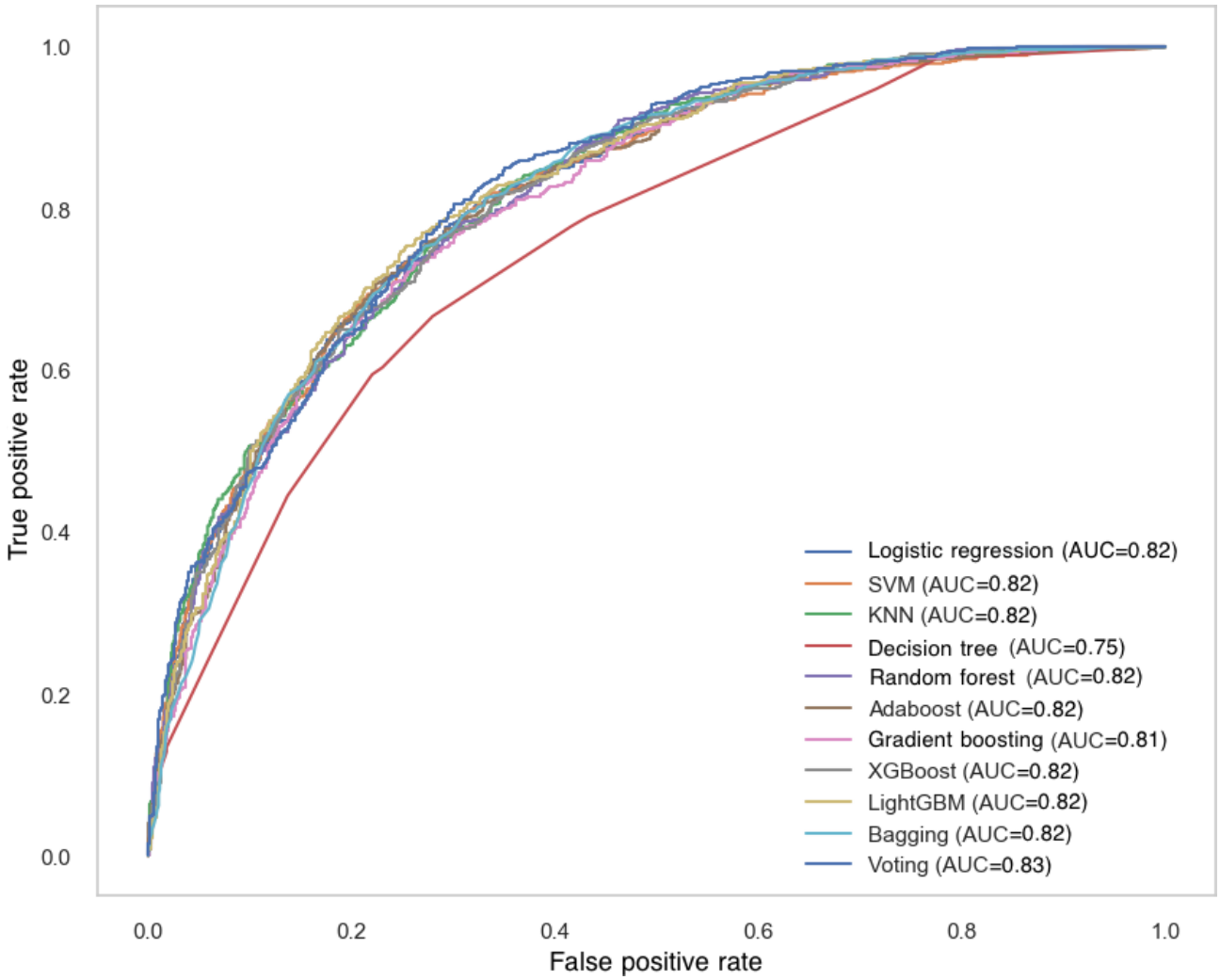
[c]KNN: k-nearest neighbors.

[d]XGBoost: extreme gradient boosting.

[e]LightGBM: light gradient boosting machine.

XSL•FO
RenderX

Figure 2. Receiver operating characteristic (ROC) curves and area under the curve (AUC) values for various classification models in predicting mortality outcomes in intensive care unit (ICU) patients with cerebrovascular diseases. KNN: k-nearest neighbors; LightGBM: light gradient boosting machine; SVM: support vector machine.



LightGBM demonstrated strong overall performance, achieving an accuracy of 74.87% and an ROC AUC of 0.82, a recall of 0.70, a precision of 0.72, and an F_1-score of 0.71. This model demonstrated robust performance across all evaluation metrics and was highly effective in discriminating between classes and minimizing false positives.

Bagging performed similarly to LightGBM, with an accuracy of 73.93%, ROC AUC of 0.82, recall of 0.72, precision of 0.70, and an F_1-score of 0.71. It also provided balanced performance but with slightly lower precision compared to LightGBM.

XGBoost and AdaBoost showed competitive performance with an accuracy of around 73.5%, both achieving an ROC AUC of 0.82. XGBoost had a recall of 0.68 and F_1-score of 0.69, while AdaBoost exhibited a slightly higher recall of 0.67 but a slightly lower precision (0.72) compared to XGBoost.

SVM and random forest both showed promising performance, with accuracy values of 74.31% and 72.92%, respectively. These models also achieved similar ROC AUC scores of 0.82, with SVM yielding a slightly higher recall of 0.66 compared to random forest (0.66). Both models displayed well-balanced performance, with SVM achieving the highest recall.

Logistic regression, a widely used baseline model for binary classification, achieved an accuracy of 74.18%, an ROC AUC of 0.82, and a recall of 0.65, demonstrating discriminative performance comparable to that of ensemble-based models such as LightGBM and Bagging.

K-nearest neighbors performed well in terms of ROC AUC (0.82) but had a relatively lower recall of 0.51, which affected its F_1-score (0.62). This highlights its ability to discriminate classes well but with a tendency to miss true positive cases.

Decision tree performed the worst, with an accuracy of 69.70%, an ROC AUC of 0.75, and relatively low precision (0.65) and recall (0.67), making it less suitable for this classification task when compared to other models.

After restructuring the experimental pipeline to ensure that imputation and feature selection were performed exclusively on the training data, model performance remained largely stable. The LightGBM model achieved an ROC AUC of 0.81 on the independent test set, indicating that the original performance estimates were not substantially inflated by data leakage.

XSL•FO
RenderX

In summary, ensemble models—particularly LightGBM and Bagging—demonstrated the best balance between sensitivity and specificity.

Temporal Validation Analysis

To further evaluate the robustness and temporal generalizability of the proposed models, we conducted a temporal validation using anchor_year_group to separate training and testing cohorts. Models trained on earlier admission years and tested on later years demonstrated stable discriminative performance. The LightGBM model achieved an AUC of 0.85, with an accuracy of 0.78, recall of 0.75, and F_1-score of 0.71. Logistic regression showed comparable performance (AUC=0.85), while the Bagging model demonstrated slightly lower discrimination (AUC=0.84). These results suggested that model discrimination remained stable when applied to temporally independent data; however, performance estimates in the temporal test set should be interpreted with caution due to potential under-ascertainment of out-of-hospital mortality in more recent calendar years of the MIMIC-IV database.

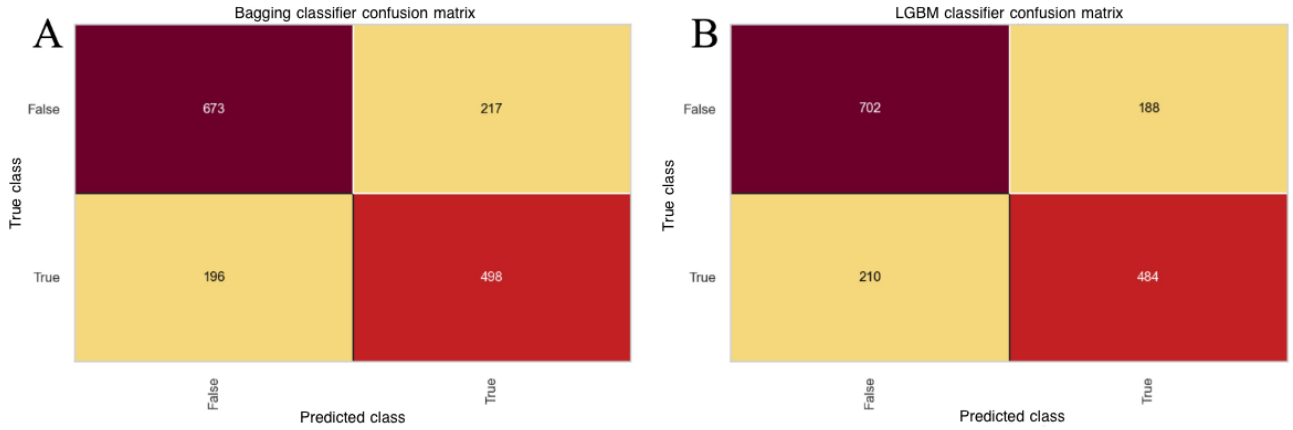
Performance Comparison With Conventional ICU Severity Scores

When evaluated as standalone predictors on the independent test set, conventional ICU severity scores demonstrated moderate discrimination. Simplified Acute Physiology Score II (SAPS-II) achieved the highest performance among traditional scores (ROC AUC=0.73; accuracy=0.65), followed by APS-III (ROC AUC=0.72; accuracy=0.66), Oxford Acute Severity of Illness Score (ROC AUC=0.66), and Logistic Organ Dysfunction System (ROC AUC=0.66). In comparison with these rule-based ICU severity scores, the LightGBM model demonstrated higher discrimination on the same test set (ROC AUC=0.83; accuracy=0.75). DeLong testing confirmed that LightGBM significantly outperformed APS-III as a standalone predictor (ΔAUC=0.12, P<.001), indicating meaningful incremental prognostic value beyond conventional ICU severity scoring systems. Notably, when compared with a fully specified multivariable logistic regression model trained on the same feature set, LightGBM showed comparable discriminative performance, highlighting that both linear and nonlinear modeling approaches can achieve similar discrimination in this setting. The added value of ensemble ML models was primarily reflected in calibration performance, DCA, and the ability to capture complex feature interactions. Importantly, this comparable performance persisted under temporal validation, indicating that both linear and nonlinear models maintained stable discrimination when applied to future patient cohorts.

Confusion Matrix and Calibration Curve

The confusion matrix for the top-performing models (LightGBM and Bagging) demonstrated high accuracy in predicting both positive (mortality) and negative (survival) outcomes (Figure 3). Bagging exhibited a strong balance between sensitivity and specificity, with a true positive rate of 0.72 and a true negative rate of 0.76.

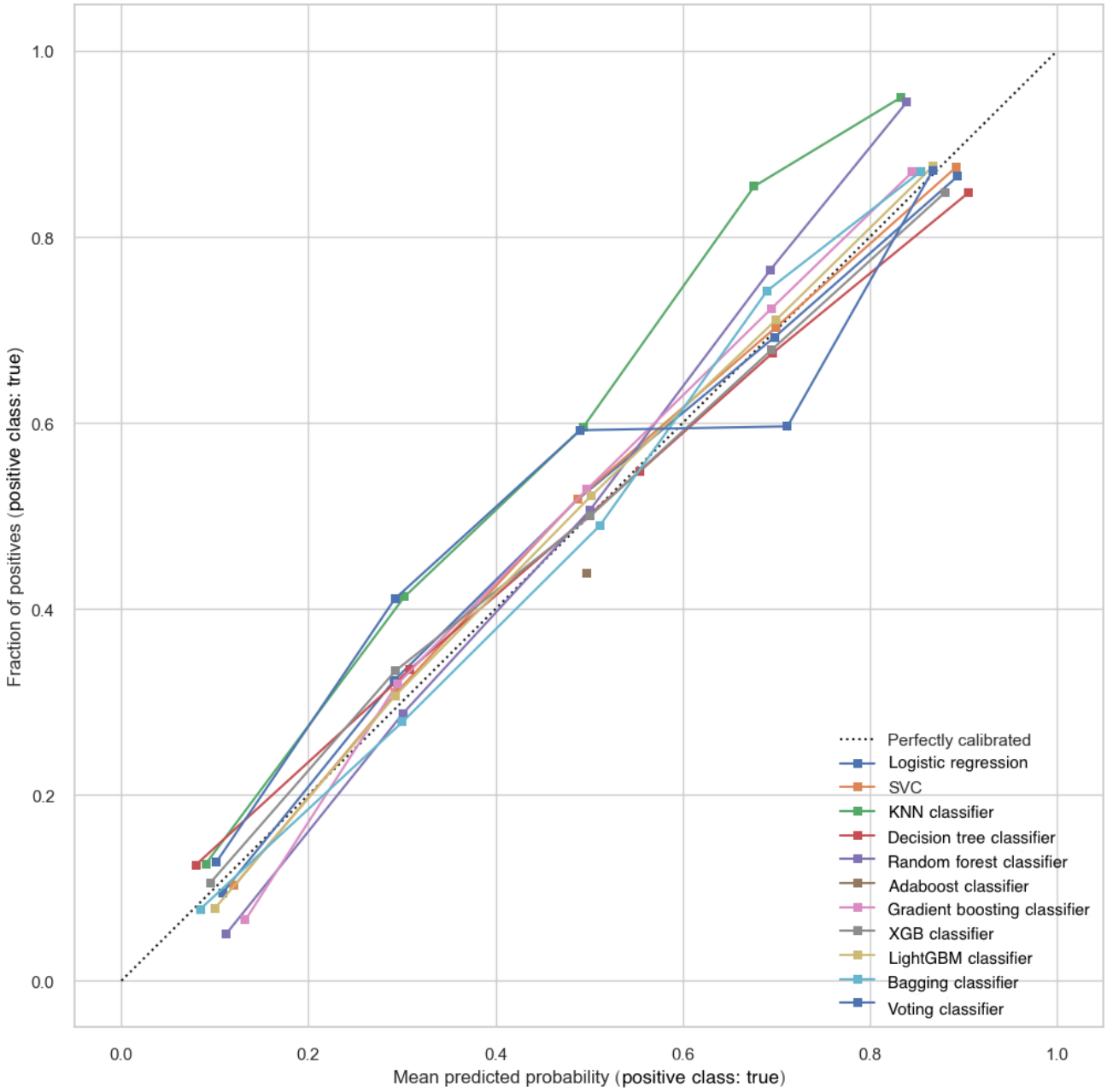
Figure 3. Confusion matrices for the Bagging and light gradient boosting machine (LightGBM) classifier models in predicting mortality outcomes in intensive care unit (ICU) patients with cerebrovascular diseases. (**A**) Confusion matrix for the Bagging classifier model. (**B**) Confusion matrix for the LightGBM classifier model. In the confusion matrices, darker red cells indicate higher counts, whereas lighter yellow cells indicate lower counts.



Calibration curves (Figure 4) were plotted to evaluate the agreement between predicted and observed mortality risk. Both LightGBM and Bagging demonstrated good calibration, with predicted probabilities closely aligning with actual outcomes across the full risk spectrum. To complement the visual assessment, quantitative calibration metrics were also computed. LightGBM achieved the lowest Brier score (0.171), indicating the best overall probability accuracy, and showed a calibration slope close to 1 (1.14) with a small intercept (0.057), suggesting minimal systematic over- or underestimation. Bagging also demonstrated strong performance with a low Brier score (0.174), a calibration slope of 1.12, and an intercept of 0.021. Together, these results indicate that LightGBM and Bagging provided well-calibrated and reliable mortality risk predictions.

XSL•FO
RenderX

Figure 4. Calibration curves for different classification models in predicting mortality outcomes in intensive care unit (ICU) patients with cerebrovascular diseases. KNN: k-nearest neighbors; LightGBM: light gradient boosting machine; SVC: support vector classifier; XGB: extreme gradient boosting.

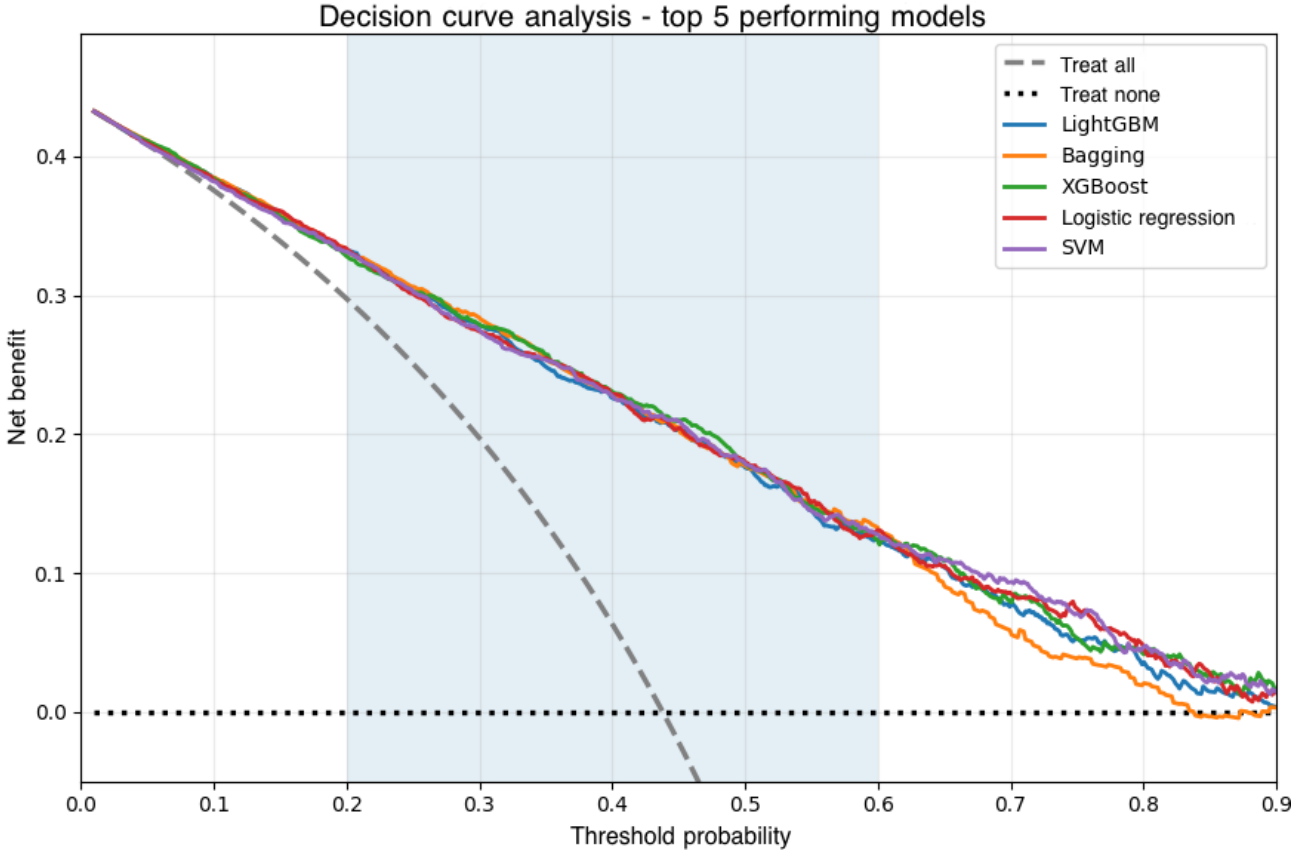


DCA and Clinical Utility

The clinical utility of the top 5 predictive models was evaluated using DCA. As shown in Figure 5, all models demonstrated a higher net benefit than the "treat none" strategy across a broad range of clinically relevant threshold probabilities. When compared with the "treat all" strategy, the ML models provided additional net benefit primarily at moderate-to-higher threshold probabilities (approximately 0.4 - 0.8), whereas at lower thresholds, their net benefit largely overlapped with the "treat all" approach.

XSL•FO
RenderX

Figure 5. Decision curve analysis (DCA) of the top five machine learning models for mortality prediction. The DCA illustrates the net benefit of using 5 predictive models (light gradient boosting machine [LightGBM], Bagging, extreme gradient boosting [XGBoost], logistic regression, and support vector machine [SVM]) across a range of threshold probabilities. The dashed gray curve represents the "treat all" strategy, which was calculated using the standard DCA definition, whereby the net benefit at a threshold probability of zero equals the observed event prevalence and decreases with increasing threshold probability. The dotted horizontal line represents the "treat none" strategy. Shaded areas indicate clinically relevant threshold probability ranges. A model with a higher net benefit at a given threshold is considered to have greater potential clinical utility.



Across this clinically relevant range, LightGBM generally showed the most favorable net benefit, with Bagging and XGBoost exhibiting closely comparable performance. Logistic regression and SVM also demonstrated consistent net benefit above the baseline strategies. Overall, these findings indicate that the proposed models may be most useful for guiding more selective clinical decisions—such as identifying high-risk patients for intensified monitoring or intervention—rather than for low-threshold scenarios where treating all patients may already be appropriate.

SHAP Analysis and Feature Importance

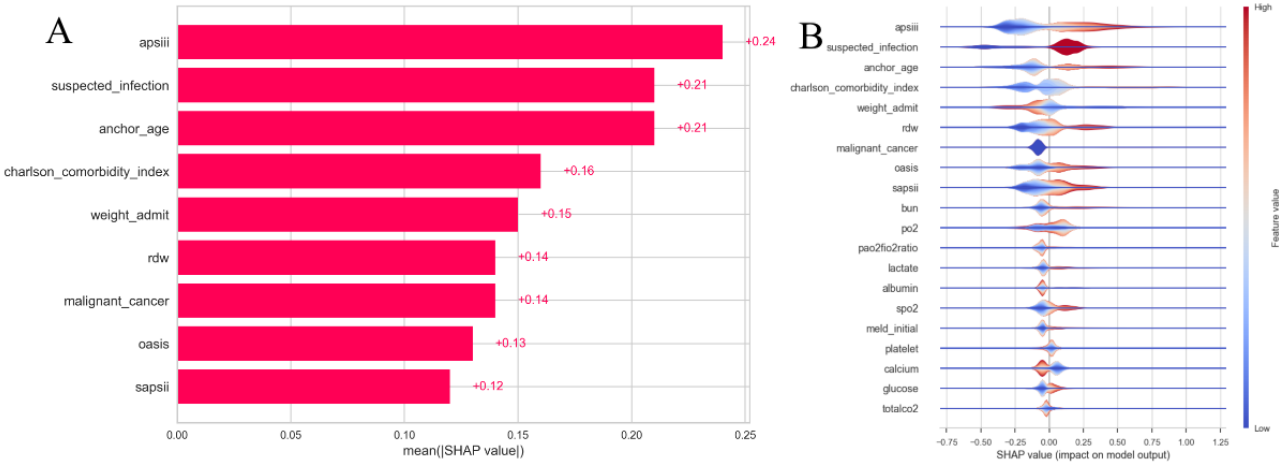
We performed SHAP analysis to interpret individual features' influence on the model's predictions. SHAP values provide a quantitative way to understand how much each feature contributes to the model's predictions. We used 2 complementary visualizations—a SHAP bar plot and a layered violin plot—to explore global feature importance and the detailed impact of each feature.

Key Insights From SHAP Analysis

The SHAP bar plot (Figure 6A) and layered violin plot (Figure 6B) both highlight the most influential features in the model. The bar plot shows the average absolute SHAP values for the top features, while the violin plot reveals the distribution of SHAP values for each feature, providing a more nuanced view of how each feature impacts the model's predictions. For clarity of visualization, the aggregated "sum of remaining features" term was excluded from the SHAP bar plot, allowing direct comparison of the relative importance of individual predictors.

XSL•FO
RenderX

Figure 6. Shapley additive explanation (SHAP) plots illustrating the impact of key features on model predictions. (**A**) Bar plot showing the average absolute SHAP values for the most influential features. (**B**) Layered violin plot displaying the distribution of SHAP values for the top features, with color indicating feature values (red for high, blue for low).



The top 5 features with the highest average SHAP values are as follows (key observations):

- “apsiii” (acute physiology score III) is the most important feature, with an average SHAP value of 0.24 (SD 0.17). This feature has a broad range of SHAP values, both positive and negative, indicating its complex impact on the model’s output depending on the severity of the patient’s condition.
- “suspected_infection” and “anchor_age” both have a mean SHAP value of 0.21 (SD 0.19 and 0.15, respectively), signifying their strong positive influence on the model’s predictions. Higher values of suspected_infection (indicating suspicion of infection) and anchor_age (older age) are associated with a higher likelihood of adverse outcomes.
- “charlson_comorbidity_index” shows a more balanced SHAP value distribution (mean SHAP value of 0.16, SD 0.19), suggesting that comorbidities have a mixed influence on the model, with both positive and negative impacts depending on the values of the index.
- “weight_admit” and “rdw” both show a strong positive contribution to the model, with average SHAP values of 0.15 (SD 0.16) and 0.14 (SD 0.13), respectively. Higher values for these features correlate with a higher probability of adverse outcomes.
- “malignant_cancer” and “oasis” also demonstrate strong positive contributions, with SHAP values predominantly concentrated in the positive range.

Feature Impact Analysis

The layered violin plot (Figure 6B) provides a deeper understanding of how the distribution of feature values influences the model’s predictions. The plot shows that the SHAP values for most of the top features are concentrated in the positive range, indicating that higher values for these features generally lead to higher predicted risks. For example, “suspected_infection,” “anchor_age,” and “weight_admit” all have SHAP values that are predominantly positive, meaning that higher values of these features are associated with a higher likelihood of adverse outcomes.

However, some features, such as “apsiii” and “charlson_comorbidity_index,” show more variability in their SHAP values, with both positive and negative values present. This indicates that these features have a more complex relationship with the model’s predictions, contributing in both directions depending on the specific values.

Overall, the visualizations help us understand which features are most important in the model’s decision-making process and how the value of each feature affects the predictions.

Summary of Key Findings

This study analyzed 5280 ICU patients with cerebrovascular diseases and found that older age, higher comorbidity scores, suspected infection, and dementia were associated with higher mortality. Feature selection identified 35 significant factors, and ML models were trained using these features. LightGBM and Bagging performed best, with LightGBM achieving 74.87% accuracy and an ROC AUC of 0.82. SHAP analysis highlighted key features such as acute physiology score III, suspected infection, and age as the most influential predictors of mortality. These models offer reliable tools for predicting all-cause mortality as recorded in the MIMIC-IV database and identifying high-risk patients in critical care settings.

Discussion

Principal Findings

This study demonstrates the potential of ML models to predict all-cause mortality among ICU patients with cerebrovascular diseases, using mortality outcomes as recorded in the MIMIC-IV database. Recent studies have shown that ML approaches can enhance prognostic accuracy in stroke and cerebrovascular disease populations, outperforming traditional clinical scores and improving individualized risk stratification [11,12]. Using the extensive MIMIC-IV database, we developed and evaluated several ML algorithms, identifying LightGBM and Bagging models as top performers across multiple metrics, consistent with the findings from recent ML-based stroke outcome prediction research [13]. These findings highlight the robustness of these models and underscore the value of ML in supporting

XSL•FO
RenderX

clinical decision-making for critically ill patients with complex cerebrovascular conditions [14].

Key Findings and Clinical Implications

Critical predictors of all-cause mortality identified in this cohort included suspected infection, age, Charlson comorbidity index, and malignancy. These factors align with prior research emphasizing the influence of comorbidities and infections on stroke outcomes [15,16]. Notably, suspected infection was highly prevalent in this cohort (73.4%), reflecting the high burden of infectious complications among critically ill patients with cerebrovascular disease admitted to the ICU, emphasizing the need for the timely recognition and management of infections—especially pneumonia—which significantly worsen stroke prognosis [17,18]. Age remains a well-established prognostic factor, with older patients facing higher mortality risk [19]. The Charlson comorbidity index further confirms the elevated risk among patients with multiple comorbid conditions, such as diabetes and hypertension [20]. SHAP analysis elucidated the relative importance of these variables, providing actionable insights for clinicians during risk assessment.

The SHAP analysis highlighted APS-III, suspected infection, comorbidity burden, and age as the most influential predictors of mortality, which is consistent with established stroke prognostic literature [21]. APS-III reflects the overall physiological severity on ICU admission, and higher scores have been strongly associated with increased mortality risk in critically ill stroke patients [22]. Infection status—particularly early respiratory or systemic infections—is a well-recognized complication that substantially increases mortality risk following ischemic stroke [23]. Similarly, a higher Charlson comorbidity index captures chronic disease burden, such as cardiovascular disease, diabetes, or malignancy, all of which worsen poststroke prognosis [24]. Age is one of the most robust and widely validated predictors of mortality and functional decline after ischemic stroke. The prominence of these variables in our SHAP results reinforces the biological plausibility of the model and indicates apparent alignment with clinical knowledge, supporting the reliability and interpretability of the ML predictions [25].

It is noteworthy that “suspected infection” was defined based on clinical interventions, including antibiotic administration and body fluid culture collection, rather than solely on physiological confirmation. As such, this feature may partially reflect clinicians’ perception of illness severity and concern for systemic infection, functioning as a process-of-care proxy rather than an independent biological marker. The considerable contribution of this variable in SHAP analyses suggests that the model may be learning from real-world clinical decision-making patterns embedded in ICU workflows. This characteristic is inherent to retrospective critical care datasets and highlights the importance of cautious interpretation when incorporating intervention-derived features into predictive models.

It should be noted that APS-III is a composite severity score derived from age and multiple physiological variables [26]. In the present models, APS-III was included alongside individual components, such as anchor_age and specific physiological measurements (eg, blood urea nitrogen and heart rate). Consequently, collinearity exists between APS-III and its constituent variables, and the contribution of age and physiological severity may be distributed across both the composite score and individual features in the SHAP-based importance rankings. Therefore, the prominence of APS-III in the SHAP analysis should be interpreted as reflecting overall illness severity rather than an independent effect separate from demographic or physiological factors, and the relative importance of these variables should be considered collectively.

Although established ICU severity scores such as APS-III and SAPS-II demonstrated reasonable predictive performance, the proposed ML models demonstrated higher discrimination than individual conventional ICU severity scores, while showing comparable performance to multivariable logistic regression models trained on the same feature set. Notably, APS-III, which represents one of the most comprehensive and widely used severity scores in critical care, was identified as the most influential individual predictor in the SHAP analysis.

Despite this, the LightGBM model provided a statistically significant improvement over APS-III alone, suggesting that the ML approach integrates complementary clinical information beyond conventional scoring systems. These findings support the role of ML as an enhancement to, rather than a replacement for, traditional ICU risk stratification tools.

Our results suggest that ML models, particularly LightGBM and Bagging, offer fairly accurate, individualized mortality predictions, enabling ICU clinicians to identify high-risk patients early. This capability could facilitate better resource allocation and more targeted interventions, ultimately improving patient outcomes where timely decision-making is critical.

In addition to predictive accuracy, the practical integration of the proposed models into clinical workflows is an important consideration. Because all input features used by the LightGBM and Bagging models are routinely collected in the ICU and are available within electronic health records, the models could be embedded into existing clinical decision support systems to generate real-time mortality risk estimates. Such integration would allow clinicians to identify high-risk patients shortly after ICU admission, prioritize monitoring intensity, optimize resource allocation, and guide discussions regarding prognosis. The use of SHAP-based explanations further enhances transparency by enabling clinicians to understand the key factors driving each patient’s prediction. Future work will focus on implementing and evaluating these models in prospective clinical settings to assess usability, workflow impact, and real-world clinical benefit.

Comparison With Existing Literature

Consistent with previous studies, our findings indicate that ML models can provide incremental prognostic value beyond conventional bedside severity scores for ICU patients with stroke-related conditions [27]. In our cohort, commonly used ICU severity scores showed moderate discrimination when evaluated as standalone predictors on the same independent test set (eg, SAPS-II AUC=0.73; APS-III AUC=0.72; Oxford Acute Severity of Illness Score AUC=0.66; Logistic Organ Dysfunction System AUC=0.66). In contrast, the best-performing ML model (LightGBM) achieved substantially

XSL•FO
RenderX

higher discrimination (AUC=0.83), and DeLong testing confirmed a statistically significant improvement over APS-III alone (ΔAUC=0.12, $P<.001$). These results suggest that the ML approach integrates complementary information beyond established scoring systems, supporting enhanced early risk stratification.

Our focus on patients with intracranial artery stenosis or occlusion extends the literature by providing detailed insights into this understudied subgroup. The large, publicly available MIMIC-IV dataset enabled robust analysis of more than 5000 patients, improving generalizability. The inclusion of mortality occurring after hospital discharge, as captured in the MIMIC-IV database, and a broad set of early clinical features, including biochemical markers and physiological severity indices, provide a comprehensive assessment of mortality risk compared with many prior studies limited by smaller samples or narrower feature sets.

Beyond traditional performance metrics, DCA demonstrated the clinical utility of the proposed models under specific decision thresholds. As shown in Figure 5, LightGBM and Bagging provided higher net benefit than the "treat none" strategy across a broad range of threshold probabilities and demonstrated clear advantage over the "treat all" strategy primarily at moderate-to-high risk thresholds. This pattern is expected given the relatively high prevalence of the outcome and indicates that the models may be particularly useful for guiding more selective clinical decisions.

In this context, incorporating these models into ICU workflows could support improved resource allocation by more accurately identifying patients at higher risk who may benefit from intensified monitoring or intervention. Together with model interpretability using SHAP, these findings support the potential for targeted, real-world application of our ML approach rather than indiscriminate use across all clinical scenarios.

Importantly, the robustness of the findings was further supported by temporal validation. When models were trained on earlier admission years and evaluated on later years, performance remained stable, with no meaningful degradation in discrimination or calibration. This temporal evaluation reduces concerns regarding information leakage inherent to random splitting in long-term retrospective datasets and supports the potential real-world applicability of the proposed models in contemporary ICU populations.

Strengths and Limitations

Key strengths include the use of multiple advanced feature selection methods (gradient boosting classifier, BorutaShap, RFECV) to isolate the most relevant predictors and the integration of SHAP for model interpretability, which fosters clinical trust and facilitates implementation.

Mortality status in this study was determined solely based on the presence or absence of a recorded dod in the MIMIC-IV database. As a result, deaths occurring outside the capture scope of the database—particularly out-of-hospital deaths in more recent calendar years—may not have been recorded. Accordingly, the absence of a documented dod should not be interpreted as confirmed long-term survival. Because mortality ascertainment relied on registry-based death records rather than active longitudinal follow-up, longer-term outcomes, such as functional recovery or late mortality beyond what was captured in the database, could not be assessed.

A related observation in this cohort was the marked discrepancy between in-hospital mortality and overall recorded mortality, indicating that a substantial proportion of deaths occurred after hospital discharge. In critically ill patients with advanced age, high comorbidity burden, malignancy, or severe physiological derangement, such postdischarge mortality may reflect transitions to hospice or comfort-focused care rather than unexpected physiological deterioration alone. Because explicit indicators of goals-of-care decisions—such as do not resuscitate status or comfort measures only orders—are not consistently captured in the public MIMIC-IV database, these factors could not be directly modeled. Consequently, the predictive models may partially capture clinical trajectories associated with end-of-life decision-making in addition to underlying illness severity. This limitation is particularly relevant in the context of temporal validation, as the incomplete capture of out-of-hospital mortality in more recent calendar years may introduce outcome misclassification, potentially inflating apparent model performance when the absence of a recorded death reflects missing data rather than true survival.

Moreover, all predictors were derived exclusively from the data available within the first 24 hours of ICU admission, prior to most formal goals-of-care decisions. However, the MIMIC-IV database does not provide reliable "present on admission" indicators for stroke diagnoses. As a result, some patients may have developed ischemic stroke after ICU admission rather than being admitted primarily for acute stroke, introducing potential temporal misclassification between disease onset and feature extraction. This limitation may affect the specificity of the cohort and should be considered when interpreting the model's applicability to strictly acute stroke populations. This design suggests that the models primarily reflect early physiological severity and comorbidity burden rather than downstream treatment choices. From a clinical perspective, the early identification of patients at extremely high risk of mortality—regardless of whether death occurs following aggressive treatment or comfort-focused care—remains highly relevant for prognostic communication, care planning, and resource allocation in the ICU.

Another important limitation is the absence of direct measures of neurological stroke severity, such as the National Institutes of Health Stroke Scale (NIHSS), which is not systematically available in the MIMIC-IV database. Without the NIHSS, the model cannot distinguish between patients with mild neurological deficits and severe systemic illness versus those with large, devastating strokes. Accordingly, the model may be better interpreted as predicting mortality among critically ill patients with stroke-associated conditions ("death with stroke") rather than mortality driven exclusively by cerebrovascular injury itself ("death from stroke"). The prominence of APS-III and other systemic physiological markers in the SHAP analysis supports this interpretation and reflects the clinical reality of ICU stroke populations, in whom outcomes are often driven by multiorgan dysfunction and comorbidity burden. Therefore, the

XSL•FO
RenderX

proposed model should not be viewed as a replacement for stroke-specific prognostic tools used in acute neurological decision-making, but rather as a complementary instrument for global mortality risk stratification in critically ill patients with stroke-associated conditions.

In addition, the feature "suspected infection" was operationally defined based on clinical actions (antibiotic administration and body fluid culture collection) rather than microbiological confirmation alone. As such, this variable may partially act as a proxy for clinician concern or illness severity, introducing potential circularity whereby the model learns from care processes rather than purely independent physiological risk factors. This reflects real-world ICU practice but warrants cautious interpretation of its predictive contribution. Although race and ethnicity were available in the database and are reported in the descriptive characteristics, this study did not perform formal subgroup or fairness analyses to evaluate potential differential model performance across demographic groups. Given known disparities in critical care delivery and the use of process-of-care–derived features such as suspected infection, algorithmic bias cannot be excluded and should be addressed in future validation studies.

The final key limitation of this study is the absence of an external validation cohort. Although rigorous internal validation was performed using an independent hold-out test set, the generalizability of the model to other institutions, populations, and clinical environments remains uncertain. Moreover, all data were derived from a single academic medical center, and institutional practices regarding discharge disposition, hospice referral, and comfort-focused care may differ substantially across health systems. As a result, the model may partially reflect center-specific end-of-life care workflows, potentially limiting its applicability to settings with different palliative care practices. At the time of this study, no external ICU dataset contained sufficiently detailed information on intracranial artery stenosis or occlusion together with reliable mortality outcomes to enable external validation. Future studies should validate these findings in multicenter datasets or prospective clinical cohorts.

Future Directions

Future research should aim to integrate these ML models into clinical workflows for real-time risk stratification. The incorporation of additional data types, such as imaging and genetic information, may further improve predictive accuracy. Prospective validation in diverse clinical settings is needed to confirm generalizability and assess impact on patient outcomes. Combining ML models with established clinical scores like NIHSS and sequential organ failure assessment could enhance predictive power and support decision-making.

In conclusion, this study supports the use of ML models to provide personalized mortality risk assessments for ICU patients with cerebrovascular disease associated with intracranial artery stenosis or occlusion, with potential to improve critical care management and patient prognosis.

Conclusions

ML models, especially LightGBM and Bagging, demonstrate comparable performance in predicting all-cause mortality in ICU patients with cerebrovascular diseases involving intracranial artery stenosis or occlusion. By leveraging routinely collected clinical data and identifying key risk factors, such as suspected infection, age, and comorbidities, these models offer valuable support for early risk stratification and clinical decision-making in critical care. Further external validation and real-world implementation studies are warranted to confirm their generalizability and clinical impact.

Acknowledgments

The authors would like to thank the PhysioNet team for maintaining the MIMIC-IV database and making it publicly available for research. The authors also acknowledge the use of ChatGPT, a generative artificial intelligence tool, for language editing and grammar refinement. All scientific content, analyses, and interpretations were generated and verified solely by the authors.

Funding

This project was supported by the Science and Technology Program of Hebei (grant number 22377712D), Medical Science Research Project of the Health Commission of Hebei Province (grant number 20260120), Hebei Provincial Health Commission (grant number 20250391), and Hebei Province Traditional Chinese Medicine Hospital Administration (grant number 2025034).

Data Availability

The datasets generated and/or analyzed during this study are available from the publicly accessible Medical Information Mart for Intensive Care IV database repository. The raw data used in the analysis, as well as the processed data, can be found in Multimedia Appendix 1. Additional datasets or analysis scripts are available upon request from the corresponding author.

Authors' Contributions

Conceptualization: KZ
Data analysis: KZ
Data collection: RC
Data curation: RC
Data interpretation: CM

XSL•FO
RenderX

Investigation: CM
Methodology: KZ
Project administration: LL
Resources: XL
Statistical analysis: YY
Supervision: LL
Validation: YY
Visualization: YY
Writing – original draft: KZ
Writing – review & editing: XL, RC, XL, JY
Writing – review & editing (text revision): PL, GX
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pearson correlation matrix of the final selected features.
[CSV File, 25 KB - cardio_v10i1e82042_app1.csv]

References

1. Zhang K, Li T, Tian J, et al. Subtypes of anterior circulation large artery occlusions with acute brain ischemic stroke. Sci Rep 2020;10(1):3442. [doi: 10.1038/s41598-020-60399-3]
2. Peterson W, Ramakrishnan N, Browder K, Sanossian N, Nguyen P, Fink E. Differentiating ischemic stroke patients from healthy subjects using a large-scale, retrospective EEG database and machine learning methods. J Stroke Cerebrovasc Dis 2024 Jun;33(6):107714. [doi: 10.1016/j.jstrokecerebrovasdis.2024.107714] [Medline: 38636829]
3. Weaver M, Goodin DA, Miller HA, et al. Prediction of prolonged mechanical ventilation in the intensive care unit via machine learning: a COVID-19 perspective. Sci Rep 2024 Dec 4;14(1):30173. [doi: 10.1038/s41598-024-81980-0] [Medline: 39627490]
4. Sheikhalishahi S, Balaraman V, Osmani V. Benchmarking machine learning models on multi-centre eICU critical care dataset. PLoS One 2020;15(7):e0235424. [doi: 10.1371/journal.pone.0235424] [Medline: 32614874]
5. Schwartz L, Anteby R, Klang E, Soffer S. Stroke mortality prediction using machine learning: systematic review. J Neurol Sci 2023 Jan 15;444:120529. [doi: 10.1016/j.jns.2022.120529] [Medline: 36580703]
6. Emami S, Martínez-Muñoz G. Condensed-gradient boosting. Int J Mach Learn & Cyber 2025 Jan;16(1):687-701. [doi: 10.1007/s13042-024-02279-0]
7. Gilotra K, Swarna S, Mani R, Basem J, Dashti R. Role of artificial intelligence and machine learning in the diagnosis of cerebrovascular disease. Front Hum Neurosci 2023;17:1254417. [doi: 10.3389/fnhum.2023.1254417] [Medline: 37746051]
8. Feng J, Zhang Q, Wu F, Peng J, Li Z, Chen Z. The value of applying machine learning in predicting the time of symptom onset in stroke patients: systematic review and meta-analysis. J Med Internet Res 2023 Oct 12;25:e44895. [doi: 10.2196/44895] [Medline: 37824198]
9. Irie F, Matsumoto K, Matsuo R, et al. Predictive performance of machine learning-based models for poststroke clinical outcomes in comparison with conventional prognostic scores: multicenter, hospital-based observational study. JMIR AI 2024 Jan 11;3(1):e46840. [doi: 10.2196/46840] [Medline: 38875590]
10. Lin CH, Chen YA, Jeng JS, et al. Predicting ischemic stroke patients' prognosis changes using machine learning in a nationwide stroke registry. Med Biol Eng Comput 2024 Aug;62(8):2343-2354. [doi: 10.1007/s11517-024-03073-4] [Medline: 38575823]
11. Wang Z, Yang W, Li Z, et al. A 25-year retrospective of the use of AI for diagnosing acute stroke: systematic review. J Med Internet Res 2024 Sep 10;26:e59711. [doi: 10.2196/59711] [Medline: 39255472]
12. Wang B, Jiang B, Liu D, Zhu R. Early predictive accuracy of machine learning for hemorrhagic transformation in acute ischemic stroke: systematic review and meta-analysis. J Med Internet Res 2025 May 23;27:e71654. [doi: 10.2196/71654] [Medline: 40408765]
13. Petrović I, Njegovan S, Tomašević O, et al. Dynamic, interpretable, machine learning-based outcome prediction as a new emerging opportunity in acute ischemic stroke patient care: a proof-of-concept study. Stroke Res Treat 2025;2025:3561616. [doi: 10.1155/srat/3561616] [Medline: 40171414]
14. Zhong X, Zhao T, Lv S, et al. Development and interpretation of a machine learning risk prediction model for post-stroke depression in a Chinese population. Sci Rep 2025;15(1):28602. [doi: 10.1038/s41598-025-09322-2]
15. Vazquez S, Stadlan Z, Lapow JM, et al. Frailty and outcomes in lacunar stroke. J Stroke Cerebrovasc Dis 2023 Feb;32(2):106942. [doi: 10.1016/j.jstrokecerebrovasdis.2022.106942] [Medline: 36525849]

XSL•FO
RenderX

16. Fluck D, Fry CH, Gulli G, et al. Association of risk of malnutrition with adverse outcomes and early support on discharge in acute stroke patients without prestroke disability: a multicenter, registry-based cohort study. Nutr Clin Pract 2022 Oct;37(5):1233-1241. [doi: 10.1002/ncp.10790] [Medline: 34664741]
17. Banda KJ, Chu H, Kang XL, et al. Prevalence of dysphagia and risk of pneumonia and mortality in acute stroke patients: a meta-analysis. BMC Geriatr 2022 May 13;22(1):420. [doi: 10.1186/s12877-022-02960-5] [Medline: 35562660]
18. Badve MS, Zhou Z, Anderson CS, Hackett ML. Effectiveness and safety of antibiotics for preventing pneumonia and improving outcome after acute stroke: systematic review and meta-analysis. J Stroke Cerebrovasc Dis 2018 Nov;27(11):3137-3147. [doi: 10.1016/j.jstrokecerebrovasdis.2018.07.001] [Medline: 30093207]
19. Bouslama M, Haussen DC, Rodrigues GM, et al. Age-adjusted infarct volume cut-off points improve stroke outcome prognostication beyond modeling with age and infarct volume. J NeuroIntervent Surg 2022 Feb;14(2):122-125. [doi: 10.1136/neurintsurg-2020-017066]
20. Liu H, Song B, Jin J, et al. Length of stay, hospital costs and mortality associated with comorbidity according to the Charlson Comorbidity Index in immobile patients after ischemic stroke in China: a national study. Int J Health Policy Manag 2021 Aug;11:1780-1787. [doi: 10.34172/ijhpm.2021.79]
21. Zhang K, Liu L, Li T, et al. Collateral circulation and rabep2 polymorphisms in large artery occlusion: impacts on short- and long-term prognosis. J Am Heart Assoc 2025 Apr 15;14(8):e040032. [doi: 10.1161/JAHA.124.040032] [Medline: 40207533]
22. Zhang GG, Hao JH, Yong Q, et al. Lactate-to-albumin ratio is associated with in-hospital mortality in patients with spontaneous subarachnoid hemorrhage and a nomogram model construction. Front Neurol 2022;13:1009253. [doi: 10.3389/fneur.2022.1009253] [Medline: 36324387]
23. Elkind MSV, Boehme AK, Smith CJ, Meisel A, Buckwalter MS. Infection as a stroke risk factor and determinant of outcome after stroke. Stroke 2020 Oct;51(10):3156-3168. [doi: 10.1161/STROKEAHA.120.030429] [Medline: 32897811]
24. Liu H, Song B, Jin J, et al. Length of stay, hospital costs and mortality associated with comorbidity according to the Charlson comorbidity index in immobile patients after ischemic stroke in China: a national study. Int J Health Policy Manag 2022 Sep 1;11(9):1780-1787. [doi: 10.34172/ijhpm.2021.79] [Medline: 34380205]
25. Chen H, Yang F, Duan Y, Yang L, Li J. A novel higher performance nomogram based on explainable machine learning for predicting mortality risk in stroke patients within 30 days based on clinical features on the first day ICU admission. BMC Med Inform Decis Mak 2024 Jun 7;24(1):161. [doi: 10.1186/s12911-024-02547-7] [Medline: 38849903]
26. Xie W, Li Y, Meng X, Zhao M. Machine learning prediction models and nomogram to predict the risk of in-hospital death for severe DKA: a clinical study based on MIMIC-IV, eICU databases, and a college hospital ICU. Int J Med Inform 2023 Jun;174:105049. [doi: 10.1016/j.ijmedinf.2023.105049] [Medline: 37001474]
27. Zhu E, Chen Z, Ai P, et al. Analyzing and predicting the risk of death in stroke patients using machine learning. Front Neurol 2023;14:1096153. [doi: 10.3389/fneur.2023.1096153] [Medline: 36816575]

Abbreviations

AdaBoost: adaptive boosting
DCA: decision curve analysis
dod: date of death
***ICD-9/10*:** *International Classification of Diseases, Ninth and Tenth Revision*
ICU: intensive care unit
LightGBM: light gradient boosting machine
MIMIC-IV: Medical Information Mart for Intensive Care IV
ML: machine learning
NIHSS: National Institutes of Health Stroke Scale
RFECV: recursive feature elimination with cross-validation
ROC AUC: area under the receiver operating characteristic curve
SAPS II: Simplified Acute Physiology Score II
SHAP: Shapley additive explanation
SVM: support vector machine
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by A Coristine; submitted 07.Aug.2025; peer-reviewed by E Bai, M Al-Agil; revised version received 27.Jan.2026; accepted 28.Jan.2026; published 24.Feb.2026.

Please cite as:
Zhang K, Chen R, Yang J, Yan Y, Liu L, Meng C, Li P, Xing G, Liu X
Machine Learning Models for Mortality Prediction in Intensive Care Unit Patients With Ischemic Stroke Associated With Intracranial Artery Stenosis: Retrospective Cohort Study
JMIR Cardio 2026;10:e82042
URL: https://cardio.jmir.org/2026/1/e82042
doi:10.2196/82042

© Kun Zhang, Ruomeng Chen, Jingyi Yang, Yan Yan, Lijuan Liu, Chaoyue Meng, Peifang Li, Guoying Xing, Xiaoyun Liu. Originally published in JMIR Cardio (https://cardio.jmir.org), 24.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Sleep and Cardiovascular Health Among Women With a History of Hypertensive Disorders of Pregnancy: Pilot Observational Study

Anais Hausvater[1], MS, MD; Elianna Shwayder[2], MD; Marie-Pierre St-Onge[3], PhD; Linda G Kahn[4,5], PhD; Manuela Plazas-Montana[1], BSc; Lillian Na[1], MS; Amanda Joa[1], BSc; Leonardo Trasande[4,5], MPP, MD; Harmony R Reynolds[1], MD

[1]Sarah Ross Soter Center for Women's Cardiovascular Research, Cardiovascular Clinical Research Center, Leon H Charney Division of Cardiology, Department of Medicine, NYU Grossman School of Medicine, 550 First Ave, New York, NY, United States

[2]Department of Medicine, Brigham and Women's Hospital, Boston, NY, United States

[3]Division of General Medicine and Center of Excellence for Sleep and Circadian Research, Columbia University Irving Medical Center, New York, NY, United States

[4]Department of Pediatrics, New York University Grossman School of Medicine, New York, NY, United States

[5]Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Anais Hausvater, MS, MD
Sarah Ross Soter Center for Women's Cardiovascular Research, Cardiovascular Clinical Research Center, Leon H Charney Division of Cardiology, Department of Medicine, NYU Grossman School of Medicine, 550 First Ave, New York, NY, United States

Abstract

Background: Both poor sleep health and hypertensive disorders of pregnancy (HDP) are independent risk factors for cardiovascular disease. Whether poor postpartum sleep contributes to the relationship between HDP and future cardiovascular disease is unknown. This pilot study evaluated the feasibility and acceptability of studying sleep health using a wearable device (Oura ring) among mothers of young children.

Objective: We evaluated indices of sleep health both objectively with the Oura ring and subjectively via questionnaires and qualitative interviews among mothers with and without a history of HDP. We also aimed to compare cardiovascular health (CVH) among mothers with vs without a prior history of HDP.

Methods: Women who were 3 to 7 years after childbirth completed baseline questionnaires (the Pittsburgh Sleep Quality Index [PSQI], Mediterranean Eating Pattern for Americans tool, and 7-item International Physical Activity Questionnaire) and wore the Oura ring continuously for 2 weeks to monitor sleep. Optimal sleep health was defined as a sleep duration of ≥7 hours, a PSQI score of ≤5, sleep timing with a sleep midpoint between 2 AM and 4 AM, a sleep efficiency of >85%, and a sleep onset variability of <60 minutes. CVH was assessed using the Life's Essential 8 score, with 8 factors assessed via questionnaires (diet, physical activity, and nicotine exposure) and objective measurements (BMI, blood pressure, blood lipids, blood glucose, and sleep duration). Semistructured interviews were conducted.

Results: In total, among 49 women, 28 (57%) with prior HDP and 21 (43%) with prior normotensive pregnancy were included, with an average of 4.9 (SD 1.2) years after delivery. Average sleep quality was suboptimal in both groups (mean PSQI score 7.0, SD 3.5 in the HDP group vs mean PSQI score 5.9, SD 2.4 in the control group; *P*=.22). Average sleep duration was suboptimal (6.7, SD 0.8 hours), with no difference between groups. Approximately half (n=23, 47%) had abnormal sleep timing, which was more common among those with a prior normotensive pregnancy. Sleep onset variability was high (mean 1.2, SD 0.5 hours), with no significant differences by HDP status. The mean CVH score fell within the moderate range (70.7, SD 12.8), with no differences between groups. The components of the CVH score that were lowest (ie, worst) among the entire cohort were diet (mean 37.3, SD 25.6) and BMI (mean 50.8, SD 35.4 kg/m^2). Common barriers to sleep included parenting, work, and household responsibilities. The study met our criteria for the feasibility and acceptability of using the Oura ring to study sleep in this population.

Conclusions: Among postpartum women, sleep health was suboptimal regardless of HDP history. Interventions to improve sleep and CVH should target all mothers during the first decade after childbirth.

(JMIR Cardio 2026;10:e81118) doi:10.2196/81118

KEYWORDS

sleep health; sleep quality; postpartum; hypertensive disorders of pregnancy; digital health; cardiovascular health

XSL•FO
RenderX

Introduction

Hypertensive disorders of pregnancy (HDP), including preeclampsia and gestational hypertension, complicate up to 15% of all pregnancies and are associated with significant maternal morbidity and mortality from cardiovascular disease (CVD) at the time of pregnancy and years after pregnancy [1-4]. Patients with a history of HDP are more likely to develop cardiovascular risk factors in the decade after delivery, including persistent hypertension and metabolic syndrome [5,6]. HDP are also associated with more than twice the risk of premature atherosclerosis and cardiovascular death, independent of other cardiovascular risk factors [7-9]. The concept of cardiovascular health (CVH) refers to a combination of modifiable factors (such as blood pressure [BP], diet, and physical activity) that, when optimized, contribute to a reduced risk of CVD. Women with prior HDP have worse CVH in the years after pregnancy than women with prior normotensive pregnancies, which is associated with an increased likelihood of subclinical CVD [10].

Sleep health is a multidimensional concept that includes several sleep indices such as duration (total sleep), efficiency (time spent asleep while in bed), timing, regularity, and quality (subjective assessment of satisfaction) [11,12]. Poor sleep health is common in US adults, is associated with cardiometabolic derangements (eg, increased adiposity [13], inflammation, and endothelial dysfunction), and predicts future cardiovascular events [13-18]. In 2022, adequate sleep duration of 7 to 9 hours per night was recognized by the American Heart Association (AHA) as a critical factor for CVH and was added to its aggregate measure of CVH, Life's Essential 8 (LE8) [16,19]. Accordingly, women are particularly susceptible to poor CVH, as they report more sleep difficulties and insomnia symptoms than men, and poor sleep is more strongly associated with CVD among women than among men [20-26]. Mothers of young children are especially at risk of poor sleep, even beyond the immediate postpartum period [27]. Short sleep during pregnancy that persists for years after childbirth was associated with an increased risk of incident metabolic syndrome [28].

Poor sleep health during pregnancy is an independent risk factor for the development of HDP [29,30]. In the years after pregnancy, small studies have shown that women with prior HDP report lower sleep quality and shorter sleep duration [31,32]. Poor sleep was more common among those with prior HDP and persistent hypertension than among those in whose BP had normalized after childbirth [33]. Women with prior preeclampsia reported poorer sleep quality and frequent sleep disturbances than women with healthy pregnancy controls 1 - 5 years after childbirth, which were associated with higher 24-hour BP [34].

The Oura ring (version Gen3; Oura Health Inc) is a wearable device that uses infrared photoplethysmography, a temperature sensor, and a 3D accelerometer to measure data such as heart rate, respiration, activity, movement, and body temperature. The ring has been successfully used in prior studies focused on women's health but has never been specifically studied among postpartum women with prior HDP [35,36]. The Oura ring is paired with a mobile app that provides users with daily feedback related to sleep health. Validation studies comparing the Oura ring with gold-standard polysomnography and research-grade actigraphy are small but have shown acceptable correlation for sleep duration and architecture [37-39]. Given the widespread commercial use of wearable devices such as the Oura ring, it is essential that we better understand their utility and potential applications in future interventions aimed at optimizing health.

The goal of this pilot study was to evaluate the use of a novel and commercially available wearable device, the Oura ring, to measure sleep health among mothers with and without a history of HDP and to better understand subjective barriers and facilitators to sleep health in this population. We also aimed to compare CVH among mothers with vs without a prior history of HDP and to evaluate how sleep health relates to the other components of CVH.

Methods

Study Participants

We performed a nested cohort pilot study examining sleep characteristics among women who were 3 to 7 years after childbirth. We recruited participants enrolled in the New York University (NYU) Children's Health and Environment Study (CHES). The methods of CHES have been previously described [40]. In brief, CHES is a prospective cohort study that recruits pregnant women aged ≥18 years who are at <18 weeks of gestation from prenatal visits at obstetrics clinics at NYU Langone-Manhattan, NYU Langone-Brooklyn, and Bellevue Hospital. The postnatal phase of CHES includes follow-up of mothers during their children's scheduled study visits (see Multimedia Appendix 1 for interview guide).

This study included CHES participants with scheduled 3 to 5 and 5 to 7 years of follow-up study visits. To account for expected sleep disturbances in the perinatal and postpartum periods, we excluded women who were currently pregnant or within 1 year after childbirth. Our cases included women with a history of HDP, defined as preeclampsia, superimposed preeclampsia, or gestational hypertension in any prior pregnancy. HDP was ascertained by manual chart review and diagnosis codes and was confirmed by participant self-report. Controls were CHES participants with a history of uncomplicated, normotensive pregnancies free of complications such as gestational diabetes or preterm birth who were attending scheduled follow-up study visits during the same time frame as cases.

During study visits, participants had their weight and BP measured and were given an Oura ring, a wearable sleep-tracking device, which they were asked to wear continuously for 2 weeks. Participants had access to the Oura mobile phone app, which provides daily statistics on sleep duration, efficiency, and timing. Participants also completed diet, physical activity, and sleep questionnaires, including the Pittsburgh Sleep Quality Index (PSQI), Mediterranean Eating Pattern for Americans tool, and the 7-item International Physical Activity Questionnaire short form, and reported nicotine exposure. Blood samples were taken from participants

XSL•FO
RenderX

(nonfasting) and analyzed for blood lipids and hemoglobin A_{1c} (a measure of blood glucose).

After the 2-week study period, select participants were invited to complete semistructured interviews via video conference. Purposive sampling was used to select participants with diverse races or ethnicities, languages, and sleep scores. They were asked questions related to barriers and facilitators to sleep health, as well as components of a sleep intervention that they thought might be feasible.

CVH Score

Participant CVH was assessed using the AHA LE8 scoring system, which includes the following factors: diet, physical activity, nicotine exposure, BMI, BP, blood lipids, blood glucose (hemoglobin A_{1c}), and sleep duration [16]. Each factor is scored from 0 to 100 based on the AHA scoring system, and the total score is an unweighted average. CVH is categorized as high (80-100), moderate (50-79), or low (0-49) [16]. Physical activity data collected using the 7-item International Physical Activity Questionnaire short form were converted to minutes per week of moderate or vigorous physical activity and mapped to the AHA LE8 scoring algorithm to derive a physical activity score from 0 to 100 [16]. If any components of the CVH score were missing, the overall score was averaged based on the number of components available (see Multimedia Appendix 2 for study questionnaires).

Sleep Health

Sleep health was measured objectively via the Oura ring wearable device and subjectively using the PSQI. Poor sleep was assessed based on 5 metrics: sleep duration (short sleep defined as <7 hours), sleep efficiency (low defined as <85%), sleep timing (suboptimal defined as a sleep midpoint before 2 AM or after 4 AM), sleep variability (elevated defined as an SD of nightly bedtime >60 minutes), and sleep quality (poor quality defined as PSQI >5) [11].

Statistical Analyses

Categorical variables, including demographic characteristics, medical history, and binary sleep measures (eg, abnormal sleep timing), were compared using the Fisher exact test. Continuous variables, including age, parity, CVH scores and their components, and continuous sleep metrics, were compared using Student *t* tests or Mann-Whitney *U* tests, as appropriate based on distribution. To evaluate whether age or current hypertension status explained observed between-group differences, we fit 3 separate regression models for each sleep outcome: an unadjusted model including group (HDP vs control), a model adjusted for age, and a model adjusted for current hypertension status, using linear models for continuous sleep outcomes and logistic regression for abnormal sleep timing. We performed Spearman correlation testing between sleep duration and the other 7 CVH component scores. Feasibility of using the Oura ring sleep tracker in this population was measured based on compliance with the Oura ring (percentage of hours worn over the 2-week period). Our feasibility target was met if >85% of participants wore the Oura ring for >70% of nights. Acceptability was measured using the Acceptability of Intervention Measure (AIM), a 4-item measure of perceived intervention acceptability. AIM items are measured on a 5-point Likert scale (completely disagree–completely agree), and the score for each item is calculated as the mean [41]. There is no universal cutoff for AIM scores. Higher AIM scores indicate a more acceptable intervention. On the basis of prior literature, we defined acceptability as an average score of ≥4 on each AIM item. Qualitative interviews were analyzed using thematic content analysis and coded using ATLAS.ti (version 25; Lumivero LLC) software. ES coded the interviews, and emergent themes were thoroughly discussed with AH.

Ethical Considerations

The study was approved by the NYU Grossman School of Medicine Institutional Review Board (i23-00046), and all participants provided written informed consent before participation. All study data were deidentified to protect the privacy and confidentiality of participants. All participants were compensated with a US $75 Amazon gift card for participation in the study.

Results

Participant Characteristics

In total, 49 participants completed the study. Of these, (n=28, 57%) had a prior delivery complicated by HDP and 21 (43%) had a prior uncomplicated pregnancy (Table 1). Assessments occurred an average of 4.9 (SD 1.2) years after pregnancy. This was a diverse cohort, with 29% (n=14) of women identified as White, 14% (n=7) as non-Hispanic Black, and 67% (n=33) as Hispanic. Women with prior HDP were older and more likely to have chronic hypertension (Table 1). There were otherwise no baseline differences in medical comorbidities between groups. Of the HDP group with 28 women, 7 (25%) had pre-existing hypertension with superimposed preeclampsia, 18 (64%) had preeclampsia, and 3 (11%) had gestational hypertension.

XSL•FO
RenderX

Table . Baseline characteristics.

Variables	All participants (n=49)	Participants with a history of hypertensive disorders of pregnancy (n=28)	Participants with a history of normotensive pregnancy (n=21)	*P* value
Age (years), mean (SD)	35.5 (6.9)	37.2 (6.8)	33.1 (6.4)	.04
Race, n (%)				.37
Asian	4 (8)	4 (14)	0 (0)	
Black or African American	7 (14)	4 (14)	3 (14)	
Other	23 (47)	12 (43)	11 (52)	
Unknown	1 (2)	0 (0)	1 (5)	
White	14 (29)	8 (29)	6 (29)	
Ethnicity, n (%)				.66
Hispanic or Latinx	33 (67)	17 (60)	16 (76)	
Not Hispanic or Latinx	15 (31)	10 (36)	5 (24)	
Unknown	1 (2)	1 (4)	0 (0)	
Marital status, n (%)				.24
Married or living with a partner	42 (86)	17 (60)	16 (76)	
Divorced or separated	2 (4)	10 (36)	5 (24)	
Single or widowed	5 (10)	1 (4)	0 (0)	
Parity, median (IQR)	2.0 (2.0-3.0)	2.0 (2.0-3.0)	2.0 (1.8-3.0)	.43
Education, n (%)				.36
High school or less	24 (49)	12 (43)	12 (57)	
Some college but no degree	7 (14)	3 (11)	4 (19)	
Associate degree	2 (4)	1 (4)	1 (4)	
Bachelor's degree	8 (16)	6 (21)	2 (10)	
Postgraduate degree	8 (16)	6 (21)	2 (10)	
Insurance, n (%)				.02
Public	36 (73)	17 (61)	19 (90)	
Private	13 (27)	11 (39)	2 (10)	
Medical history, n (%)				
Obesity (BMI ≥30 kg/m^2)	25 (51)	13 (46)	12 (57)	.57
Hypertension	14 (29)	13 (46)	1 (5)	<.001
Diabetes	4 (8)	3 (11)	1 (5)	.68
High cholesterol	5 (10)	2 (7)	3 (14)	.64
Obstructive sleep apnea	2 (4)	1 (4)	1 (5)	1

Feasibility and Acceptability

We consented 52 participants and accrued 49 participants for this study over 12 months. Overall, 3 women did not initiate the study due to technological challenges. All included participants had a minimum of 7 nights of data. Of the 49 participants, 90% (n=44) wore the Oura ring for >70% of nights over the 2-week study period, indicating excellent adherence. Our acceptability target was met (median AIM item score 4, IQR 4-5; Table S1 in Multimedia Appendix 3). Participants' feedback on the ring-wearing experience was generally positive, including: "It (Oura ring) is very easy to use," "(Oura ring is) a great device," and "It (Oura app) helped me realize that I don't sleep well."

Sleep Health

In the unadjusted analysis, there were no differences in the proportion of women who achieved optimal sleep quality,

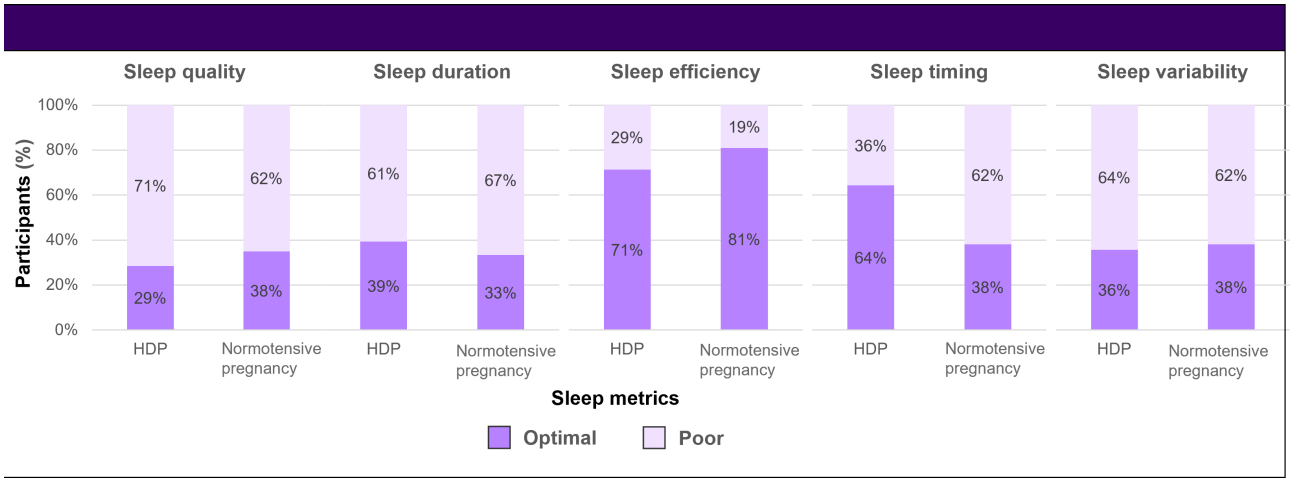
XSL•FO
RenderX

efficiency, duration, timing, or regularity when comparing those with vs without prior HDP (Table 2, Figure 1). Sleep quality was suboptimal on average across the cohort and was similar among HDP and control participants (mean PSQI score 7.0, SD 3.5 in the HDP group vs mean PSQI score 5.9, SD 2.4 in the control group; P=.22) [Table 2]. Oura ring-derived sleep data revealed that average sleep duration was suboptimal for most mothers (mean 6.7, SD 0.8 hours), with no difference between HDP and control participants (P=.99). Sleep timing was abnormal in 47% of participants overall. Sleep onset variability was high on average across the entire cohort, reflecting suboptimal regularity in bedtimes (mean 1.2, SD 0.5 hours), but was similar between groups (P=.97). Sleep efficiency was normal overall and did not differ between groups (mean 86.9%, SD 3.8%; P=.62) [Table 2]. Approximately one-third of participants met criteria for optimal sleep across individual domains.

Table . Sleep outcomes.

Sleep outcomes	All participants (n=49)	Participants with a history of hypertensive disorders of pregnancy (n=28)	Participants with normotensive pregnancy (n=21)
Sleep duration (hours), mean (SD)	6.65 (0.80)	6.65 (0.75)	6.66 (0.83)
Sleep efficiency (%), mean (SD)	86.9 (3.83)	86.65 (4.27)	87.21 (3.24)
Abnormal sleep timing, n (%)	23 (47)	10 (36)	13 (62)
Sleep variability (hours), mean (SD)	1.22 (0.52)	1.22 (0.60)	1.21 (0.41)
Sleep quality (Pittsburgh Sleep Quality Index), mean (SD)	6.49 (3.09)	6.96 (3.48)	5.86 (2.41)

Figure 1. Sleep metrics among women with and without a history of hypertensive disorders of pregnancy (HDP). Optimal sleep quality was defined as a Pittsburgh Sleep Quality Index score of ≤5. Optimal sleep duration, efficiency, timing, and variability were derived from Oura ring data and defined as follows: duration ≥7 hours per night, sleep efficiency ≥85%, sleep timing with a sleep midpoint between 2 AM and 4 AM, and sleep variability defined as an SD in bedtime <1 hour.



In regression analyses, HDP was not associated with sleep duration, sleep efficiency, sleep quality (PSQI), continuous sleep timing, or sleep variability in either unadjusted or age-adjusted models; findings were similarly nonsignificant in models adjusted for baseline hypertension (Table 3). After adjustment for current hypertension status, abnormal sleep timing—defined as a sleep midpoint outside the interval of 2 AM to 4 AM—was more common among those with prior normotensive pregnancy (Table 3). The corresponding unadjusted and age-adjusted estimates showed a trend toward more abnormal sleep timing among those with prior normotensive pregnancy (both P≈.07).

XSL•FO
RenderX

Table . Linear regression estimates comparing sleep metrics between participants with a history of hypertensive disorders of pregnancy (HDP) and participants with normotensive pregnancy (control).[a]

Sleep metrics	β coefficient[b]	95% CI	*P* value
Sleep duration			
Unadjusted	0	−0.46 to 0.46	.99
Adjusted for age	0.15	−0.32 to 0.61	.53
Adjusted for hypertension	0.22	−0.28 to 0.72	.39
Sleep efficiency			
Unadjusted	−0.56	−2.80 to 1.68	.62
Adjusted for age	0.31	−1.93 to 2.55	.78
Adjusted for hypertension	0.74	−1.66 to 3.15	.54
Abnormal sleep timing (continuous)			
Unadjusted	−0.66	−1.45 to 0.14	.10
Adjusted for age	−0.53	−1.37 to 0.31	.21
Adjusted for hypertension	−0.69	−1.59 to 0.22	.13
Sleep variability			
Unadjusted	0.01	−0.30 to 0.31	.97
Adjusted for age	−0.03	−0.35 to 0.30	.88
Adjusted for hypertension	−0.16	−0.49 to 0.17	.32
Sleep quality (Pittsburgh Sleep Quality Index score)			
Unadjusted	1.11	−0.68 to 2.89	.22
Adjusted for age	1.28	−0.61 to 3.17	.40
Adjusted for hypertension	0.33	−1.63 to 2.30	.11
Abnormal sleep timing (binary)			
Unadjusted	−1.07	−2.28 to 0.08	.07
Adjusted for age	−1.14	−2.44 to 0.08	.07
Adjusted for hypertension	−1.52	−3.04 to −0.18	.03

[a]Values are presented as β coefficients representing the adjusted mean difference (HDP − control) for linear regression and log-odds for logistic regression, with 95% CIs and *P* values shown for unadjusted models and models adjusted for age or current hypertension.

[b]β coefficients represent mean differences for all outcomes except abnormal sleep timing (binary), for which β coefficients represent log-odds.

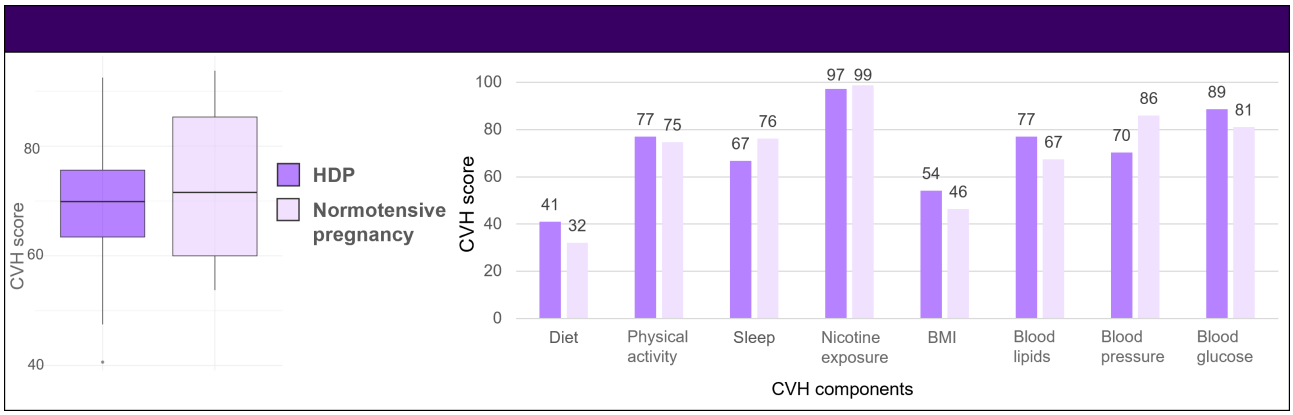
Optimal sleep quality was defined as a PSQI score of ≤5. Optimal sleep duration, efficiency, timing, and variability were derived from Oura ring data and defined as follows: sleep duration ≥7 hours per night, sleep efficiency ≥85%, sleep timing with a sleep midpoint between 2 AM and 4 AM, and sleep variability defined as an SD in bedtime <1 hour.

CVH Score Analysis

AHA CVH score was in the moderate range overall (mean 70.77, SD 12.88), with no difference between groups (mean 70.9, SD 11.4 vs mean 70.3, SD 14.7; *P*=.87; Figure 2, Table S2 in Multimedia Appendix 3). The components of the CVH score that were lowest (ie, worst) among the entire cohort were diet (mean 37.3, SD 25.6) and BMI scores (50.8, SD 35.4; Table S2 in Multimedia Appendix 3). There were no differences in any components of the CVH score between those with vs without prior HDP (Figure 2, Table S2 in Multimedia Appendix 3), but there was a trend toward lower scores (ie, higher BPs) among the group with prior HDP that was not statistically significant (*P*=.07; Table S2 in Multimedia Appendix 3). In the full cohort, there were no significant correlations between the CVH component score for sleep health and the other 7 components of LE8 (Table S3 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Cardiovascular health (CVH) scores among women with and without a history of hypertensive disorders of pregnancy (HDP). The CVH score reflects the unweighted average of 8 components: diet, physical activity, sleep, nicotine exposure, BMI, blood lipids, blood pressure, and blood glucose. Each component was scored from 0 to 100 based on the American Heart Association Life's Essential 8 criteria. Higher scores indicate better health.



Patient Perspectives

In total, 19 (39%) participants were invited to complete qualitative interviews (PSQI scores range 1-14; n=19, 42% Hispanic). Of these, 14 (74%) participants consented and were interviewed about their experiences after the study period (PSQI scores 1-14; n=14, 36% Hispanic). Of the 14 participants, 12 (86%) had prior HDP. All participants noted that tracking their sleep metrics had some influence on them, ranging from increased awareness of poor sleep habits to active modification of behaviors related to bedtime, sleep duration, or sleep hygiene.

Only 2 (14.3%) participants described a dedicated bedtime routine including habits explicitly intended to facilitate sleep, such as avoiding screens, maintaining a consistent bedtime, drinking tea, taking a bath, and reading. Overall, 6 (42.9%) participants identified sleep-facilitating practices that they engaged in more sporadically, such as meditating, reading, drinking tea, and using scented oils. Only 3 (21.4%) participants reported routines centered around their children's bedtime, including reading, singing, and cuddling. Only 3 (21.4%) participants lacked a consistent bedtime routine. Overall, 11 (78.6%) participants described behaviors and responsibilities that adversely impacted sleep, including late-night screen time (n=8, 57.1%), inconsistent bedtimes (n=6, 42.9%), late-night parenting duties (n=6, 42.9%), working late (n=6, 42.9%), and household chores (n=2, 14.3%). The remaining 3 (21.4%) participants did not identify specific activities that hindered sleep.

Barriers to adequate sleep were thematically related to societal demands, family responsibilities, and individual factors. Societal demands included competing work, domestic, and caregiver responsibilities, and gender-related norms such as the burden of childcare falling on the mother (eg, child health- or school-related challenges and the "mental load" of parenting). Financial stressors and environmental disturbances (eg, city noise) were also identified as external factors that negatively impacted sleep. Themes centered on the family unit included parenting roles (eg, child-centered bedtime routines), dynamics with partners (eg, maintaining relationship with a significant other), and child-specific barriers (eg, cosleeping and child sleep difficulty). Individual barriers included a desire for personal time (often only available after children were asleep), a lack of prioritization of sleep health in general, a lack of consistency with sleep-facilitating activities, being a light sleeper, high stress levels, and waking during the night.

When prompted to consider modifiable aspects of their sleep routine, 7 (50%) participants favored bedtime consistency, 4 (28.6%) participants wanted to sleep 7 to 8 hours every night, 1 (7.1%) participant viewed both approaches as feasible, and 2 (14.3%) participants felt their sleep habits were not modifiable. When asked what type of messaging might motivate mothers to engage in a sleep intervention, 6 (42.9%) participants believed that motherhood-focused messages would be most salient ("If I had better sleep habits, I'd be more aware of my child and be a better mother"), 3 (21.4%) participants favored emphasizing personal health ("If you sleep enough, you're more productive and gain health benefits"), and 2 (14.3%) participants advocated for emphasizing the intertwined nature of personal health, parenting, and quality of life ("Improving your health means staying around longer for your kids, modeling good habits, and being less tired to care for them"). The remaining 3 (21.4%) participants did not identify a specific motivating messaging approach or expressed uncertainty about what type of messaging would influence their behavior.

Discussion

Principal Findings

In this pilot study, the Oura ring wearable device was a feasible and acceptable tool for measuring sleep in this population of mothers of young children. Mothers of children aged 3 to 7 years had suboptimal sleep quality, duration, and regularity, independent of prior HDP status. Barriers to sleep health in this population included parenting, professional, and domestic responsibilities. Most participants lacked a consistent bedtime routine and regularly engaged in behaviors that adversely affect sleep. Most participants cited that tracking their sleep with the Oura ring inspired awareness of poor sleep health. These data provide support for the development of future sleep optimization intervention studies aimed at optimizing CVH.

In regression analyses, there was a significant association between abnormal sleep timing and prior normotensive

XSL•FO
RenderX

pregnancy in the model adjusted for current hypertension status. This finding was not observed in unadjusted or age-adjusted models. Therefore, adjustment for current hypertension may represent overadjustment, as chronic hypertension could lie on the causal pathway between HDP and later sleep or CVH. As such, this result is considered exploratory and hypothesis-generating and requires confirmation in larger, adequately powered cohorts.

Prior Work

To our knowledge, this is the first study to evaluate sleep indices objectively using a wearable device in a population of mothers within the decade after delivery. Our findings are in line with prior studies showing that maternal sleep is poor years after delivery [27,42]. Insomnia and short sleep (<7 hours per night) are common among mothers in the first decade after delivery [28,42]. Sleep duration, as measured by the Oura ring, was overall worse when compared with a cohort of healthy young and midlife women from the United States [43].

Previous studies that relied solely on subjective sleep metrics found that years after pregnancy, women with prior HDP report poor sleep quality and short sleep duration [31,32]. In one such study of 526 women, 20% of whom had HDP, the PSQI was completed to evaluate their sleep quality an average of 13 years after delivery. Sleep quality scores were poor overall and were slightly worse among women with prior HDP than those with normotensive pregnancies (6.6 vs 5.8; P=.04) [31]. Similar to this study's analysis, the investigators found that sleep duration was suboptimal but did not differ between groups [31]. A small study of 13 women with prior preeclampsia reported poorer sleep quality and more frequent sleep disturbances than healthy pregnancy controls 1 to 5 years after childbirth, and these sleep problems were associated with higher 24-hour BP [34]. In this study, we found that certain metrics, such as sleep timing and some CVH metrics, including diet, BMI, and blood lipids, were numerically worse among participants with prior normotensive pregnancies. This might be explained by the small numbers of participants in each group, as well as differences in socioeconomic and cultural factors between groups. The control arm had a higher proportion of Hispanic participants, lower levels of education, and a higher proportion of public insurance. These differences in socioeconomic factors alone may lead to poor sleep and CVH. Taken together, sleep optimization strategies may benefit all mothers in the first decade after delivery and may be of particular importance among those at higher cardiovascular risk and lower socioeconomic status.

To our knowledge, ours is the first study to combine both objective and subjective sleep data in this population, with the aim of capturing the full spectrum of sleep health. Prior literature suggests that subjective measures of sleep health can overestimate or underestimate metrics such as sleep duration, efficiency, and sleep timing [44]. There is poor correlation between perceptions of sleep as "poor" or "sufficient" and objective sleep measures of sleep duration and depth [44]. Individual perceptions of sleep quality and daytime functioning are subjective measures of sleep health that provide different but equally essential information. Thus, both objective and subjective sleep data are required to fully understand the whole spectrum of a person's sleep health.

Despite the growing recognition of sleep health as an important contributor to cardiovascular risk, most physicians fail to routinely inquire about a patient's sleep health or counsel patients on how to improve sleep [45]. This pilot study demonstrates that it is feasible to track sleep health using a wearable device among mothers of young children. Importantly, sleep tracking with the Oura ring heightened participants' awareness of their suboptimal sleep health, and mothers were receptive to feedback regarding their sleep behaviors. This increased awareness underscores an unmet need for patient education on the connection between sleep health and cardiovascular risk and lifestyle measures that can improve CVH. This education may be particularly pertinent for women with prior HDP, given their limited awareness of elevated cardiovascular risk [46]. Once informed, women with a history of HDP express a strong desire for clinician guidance to support long-term lifestyle change [46]. Similarly, in our cohort, participants discussed that the sleep data they received from the Oura ring helped them identify feasible behavioral targets, namely consistent bedtime and increased sleep duration. The potential for wearable sleep trackers such as the Oura ring to facilitate durable improvements in sleep behavior remains uncertain. Digital health interventions, including wearable devices, may improve cardiovascular risk factors such as BP, physical activity, and weight among postpartum women, but studies to date have been small and have shown mixed results [39].

Strengths and Limitations

This is the first study to examine the feasibility and acceptability of using the Oura ring, a popular commercial device, to measure sleep health in mothers of young children. However, there are several limitations worth noting. First, this pilot study was not sufficiently powered to detect clinically meaningful differences in sleep health and CVH components between groups. However, these data may be used to estimate variance for powering larger observational studies. Second, we only monitored sleep for 2 weeks, which may not have captured the full spectrum of participants' sleep habits. To mitigate this limitation, participants were instructed to wear the ring during periods when they are following their habitual schedule (eg, not while on vacation). It is possible that sleep behaviors changed in response to wearing the Oura ring, but this is likely mitigated by the fact that we are reporting averages across the study period between both groups. Third, the Oura ring has been validated against research-grade actigraphy in only small-sized studies [37]. Fourth, we did not include screening for sleep apnea or insomnia, which limited our ability to differentiate sleep disorders from poor sleep health. Finally, this was a substudy of a larger observational cohort (CHES), which may have introduced selection bias.

Future Directions

Our findings suggest that sleep and CVH optimization efforts should extend beyond women with prior HDP to a broader postpartum population in the first decade after delivery. Future studies in larger cohorts are necessary to determine whether wearable sleep-tracking interventions effectively support

XSL•FO
RenderX

sustained behavioral change, particularly among mothers navigating the unique sleep-related challenges associated with early parenthood.

Conclusions

In this cohort of postpartum mothers, we found that it was feasible and acceptable to use the Oura ring to study sleep among women years after delivery. We found that sleep quality, duration, and variability were uniformly poor, with no differences between those with and without prior HDP. CVH was suboptimal across all participants.

Acknowledgments

The authors thank all the participants and staff members at the New York University Children's Health and Environment Study for their time and effort.

Funding

This research was funded by the National Center for Advancing Translational Sciences Clinical and Translational Science Institute pilot award (grant UL1TR001445), which provided US $40,000 over 2 years (January 2023 to December 2024). Dr Hausvater was supported by the National Institutes of Health (NIH)–National Heart, Lung, and Blood Institute (grant T32HL098129). New York University (NYU) Children's Health and Environment Study is supported by institutional funds from NYU Grossman School of Medicine as well as the NIH Office of the Director (grant UG/H3OD023305).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AH (lead), MPS-O, LGK, MP-M, LT, HRR
Data curation: AH, ES, LN, AJ
Formal analysis: AH (lead), ES (supporting)
Funding acquisition: AH
Investigation: AH, ES, MP-M, LN, AJ
Methodology: AH, MPS-O, LGK, MP-M
Project administration: AH (lead), MP-M (supporting)
Resources: LT
Supervision: HRR, LGK, LT
Validation: AH
Visualization: AH, ES
Writing – original draft: AH (lead), ES (supporting)
Writing – review & editing: AH (lead), MPS-O, LGK, MP-M, LN, AJ, HRR

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative interview guide.
[DOCX File, 14 KB - cardio_v10i1e81118_app1.docx]

Multimedia Appendix 2
Study questionnaires.
[PDF File, 1212 KB - cardio_v10i1e81118_app2.pdf]

Multimedia Appendix 3
Additional study tables on Acceptability of Intervention Measure questionnaire results, cardiovascular health and its components, and correlations between the sleep health score and cardiovascular health components.
[DOCX File, 18 KB - cardio_v10i1e81118_app3.docx]

References

1. Shah LM, Varma B, Nasir K, et al. Reducing disparities in adverse pregnancy outcomes in the United States. Am Heart J 2021 Dec;242:92-102. [doi: 10.1016/j.ahj.2021.08.019] [Medline: 34481757]

XSL•FO
RenderX

2. Ying W, Catov JM, Ouyang P. Hypertensive disorders of pregnancy and future maternal cardiovascular risk. J Am Heart Assoc 2018 Sep 4;7(17):e009382. [doi: 10.1161/JAHA.118.009382] [Medline: 30371154]
3. Ford ND, Cox S, Ko JY, et al. Hypertensive disorders in pregnancy and mortality at delivery hospitalization - United States, 2017-2019. MMWR Morb Mortal Wkly Rep 2022 Apr 29;71(17):585-591. [doi: 10.15585/mmwr.mm7117a1] [Medline: 35482575]
4. Cameron NA, Everitt I, Seegmiller LE, Yee LM, Grobman WA, Khan SS. Trends in the incidence of new-onset hypertensive disorders of pregnancy among rural and urban areas in the United States, 2007 to 2019. J Am Heart Assoc 2022 Jan 18;11(2):e023791. [doi: 10.1161/JAHA.121.023791] [Medline: 35014858]
5. Behrens I, Basit S, Melbye M, et al. Risk of post-pregnancy hypertension in women with a history of hypertensive disorders of pregnancy: nationwide cohort study. BMJ 2017 Jul 12;358:j3078. [doi: 10.1136/bmj.j3078] [Medline: 28701333]
6. Ehrenthal DB, McNeil RB, Crenshaw EG, et al. Adverse pregnancy outcomes and future metabolic syndrome. J Womens Health (Larchmt) 2023 Sep;32(9):932-941. [doi: 10.1089/jwh.2023.0026] [Medline: 37262199]
7. Ray JG, Vermeulen MJ, Schull MJ, Redelmeier DA. Cardiovascular health after maternal placental syndromes (CHAMPS): population-based retrospective cohort study. Lancet 2005 Nov 19;366(9499):1797-1803. [doi: 10.1016/S0140-6736(05)67726-4] [Medline: 16298217]
8. Wang YX, Arvizu M, Rich-Edwards JW, et al. Hypertensive disorders of pregnancy and subsequent risk of premature mortality. J Am Coll Cardiol 2021 Mar 16;77(10):1302-1312. [doi: 10.1016/j.jacc.2021.01.018] [Medline: 33706872]
9. Sederholm Lawesson S, Swahn E, Pihlsgård M, et al. Association between history of adverse pregnancy outcomes and coronary artery disease assessed by coronary computed tomography angiography. JAMA 2023 Feb 7;329(5):393-404. [doi: 10.1001/jama.2022.24093] [Medline: 36749333]
10. Benschop L, Schalekamp-Timmermans S, Schelling SJ, Steegers EA, Roeters van Lennep JE. Early pregnancy cardiovascular health and subclinical atherosclerosis. J Am Heart Assoc 2019 Aug 6;8(15):e011394. [doi: 10.1161/JAHA.118.011394] [Medline: 31331213]
11. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 1;37(1):9-17. [doi: 10.5665/sleep.3298] [Medline: 24470692]
12. St-Onge MP, Aggarwal B, Fernandez-Mendoza J, et al. Multidimensional sleep health: definitions and implications for cardiometabolic health: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2025 May;18(5):e000139. [doi: 10.1161/HCQ.0000000000000139] [Medline: 40223596]
13. Covassin N, Singh P, McCrady-Spitzer SK, et al. Effects of experimental sleep restriction on energy intake, energy expenditure, and visceral obesity. J Am Coll Cardiol 2022 Apr 5;79(13):1254-1265. [doi: 10.1016/j.jacc.2022.01.038] [Medline: 35361348]
14. Shah R, Shah VK, Emin M, et al. Mild sleep restriction increases endothelial oxidative stress in female persons. Sci Rep 2023 Sep 16;13(1):15360. [doi: 10.1038/s41598-023-42758-y] [Medline: 37717072]
15. Tartar JL, Fins AI, Lopez A, et al. Sleep restriction and delayed sleep associate with psychological health and biomarkers of stress and inflammation in women. Sleep Health 2015 Dec;1(4):249-256. [doi: 10.1016/j.sleh.2015.09.007] [Medline: 29073400]
16. Lloyd-Jones DM, Allen NB, Anderson CA, et al. Life's Essential 8: updating and enhancing the American Heart Association's construct of cardiovascular health: a presidential advisory from the American Heart Association. Circulation 2022 Aug 2;146(5):e18-e43. [doi: 10.1161/CIR.0000000000001078] [Medline: 35766027]
17. Makarem N, Castro-Diehl C, St-Onge MP, et al. Redefining cardiovascular health to include sleep: prospective associations with cardiovascular disease in the MESA Sleep Study. J Am Heart Assoc 2022 Nov;11(21):e025252. [doi: 10.1161/JAHA.122.025252] [Medline: 36259552]
18. Sunbul EA, Sunbul M, Tastekin N, Ozben B, Sayar N, Gulec H. Is there any relation between arterial stiffness and insomnia? A challenging question. Sleep Breath 2022 Mar;26(1):333-338. [doi: 10.1007/s11325-021-02445-6] [Medline: 34302608]
19. Isiozor NM, Kunutsor SK, Voutilainen A, Laukkanen JA. Life's Essential 8 and the risk of cardiovascular disease death and all-cause mortality in Finnish men. Eur J Prev Cardiol 2023 Jun 1;30(8):658-667. [doi: 10.1093/eurjpc/zwad040] [Medline: 36753230]
20. Meers J, Stout-Aguilar J, Nowakowski S. Chapter 3 - Sex differences in sleep health. In: Grandner MA, editor. Sleep and Health: Academic Press; 2019:21-29. [doi: 10.1016/B978-0-12-815373-4.00003-4]
21. Cappuccio FP, Stranges S, Kandala NB, et al. Gender-specific associations of short sleep duration with prevalent and incident hypertension: the Whitehall II Study. Hypertension 2007 Oct;50(4):693-700. [doi: 10.1161/HYPERTENSIONAHA.107.095471] [Medline: 17785629]
22. St-Onge MP, Grandner MA, Brown D, et al. Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. Circulation 2016 Nov 1;134(18):e367-e386. [doi: 10.1161/CIR.0000000000000444] [Medline: 27647451]
23. Laugsand LE, Vatten LJ, Platou C, Janszky I. Insomnia and the risk of acute myocardial infarction: a population study. Circulation 2011 Nov 8;124(19):2073-2081. [doi: 10.1161/CIRCULATIONAHA.111.025858] [Medline: 22025601]

XSL•FO
RenderX

24. van Egmond LT, Meth EM, Engström J, et al. Effects of acute sleep loss on leptin, ghrelin, and adiponectin in adults with healthy weight and obesity: a laboratory study. Obesity (Silver Spring) 2023 Mar;31(3):635-641. [doi: 10.1002/oby.23616] [Medline: 36404495]
25. Zuraikat FM, Jelic S, St-Onge MP. Wake up! It's time to recognize the importance of sleep in metabolic health. Obesity (Silver Spring) 2023 May;31(5):1188-1191. [doi: 10.1002/oby.23776] [Medline: 37140407]
26. Lyytikäinen P, Lallukka T, Lahelma E, Rahkonen O. Sleep problems and major weight gain: a follow-up study. Int J Obes (Lond) 2011 Jan;35(1):109-114. [doi: 10.1038/ijo.2010.113]
27. Boergers J, Hart C, Owens JA, Streisand R, Spirito A. Child sleep disorders: associations with parental sleep duration and daytime sleepiness. J Fam Psychol 2007 Mar;21(1):88-94. [doi: 10.1037/0893-3200.21.1.88] [Medline: 17371113]
28. Kim M, Wiener LE, Gilbert J, et al. Persistent short sleep duration from pregnancy to 2 to 7 years after delivery and metabolic health. JAMA Netw Open 2024 Dec 2;7(12):e2452204. [doi: 10.1001/jamanetworkopen.2024.52204] [Medline: 39724370]
29. Zhu J, Zheng T, Jin H, et al. Sleep disturbances in early gestation and the risks of hypertensive disorders of pregnancy: a prospective cohort study. Am J Epidemiol 2024 Feb 5;193(3):479-488. [doi: 10.1093/aje/kwad223] [Medline: 37968336]
30. Sanapo L, Bublitz MH, Bourjeily G. Sleep disordered breathing, a novel, modifiable risk factor for hypertensive disorders of pregnancy. Curr Hypertens Rep 2020 Mar 12;22(4):28. [doi: 10.1007/s11906-020-1035-7] [Medline: 32166454]
31. Aggarwal BA, Hauspurg A, St-Onge MP, Jelic S, Catov JM. Abstract P237: Poor sleep quality after delivery: a contributor to increased cardiovascular disease risk for women with gestational hypertension and preeclampsia? Circulation 2020 Mar;141(Suppl_1). [doi: 10.1161/circ.141.suppl_1.P237]
32. Jackson C, Gaston S, Louis J, Sandler D. 0819 Does sleep health modify the relationship between hypertensive disorders of pregnancy and incident cardiovascular disease? Sleep 2023 May;46(Supplement_1):A360-A361. [doi: 10.1093/sleep/zsad077.0819]
33. Wang T, Bryan S, Jeyabalan A, et al. Sleep quality in women with hypertensive disorders of pregnancy with and without persistent postpartum hypertension. Am J Obstet Gynecol 2022 Jan;226(1):S278. [doi: 10.1016/j.ajog.2021.11.473]
34. Nuckols V, Guerrero M, Stroud A, Santillan MK, Pierce GL. Abstract P098: association between postpartum sleep characteristics and twenty-four hour blood pressure patterns in women with and without a history of preeclampsia. Hypertension 2022 Sep;79(Suppl_1):AP098. [doi: 10.1161/hyp.79.suppl_1.P098]
35. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable Oura ring. BMC Womens Health 2019 Nov 29;19(1):150. [doi: 10.1186/s12905-019-0844-9] [Medline: 31783840]
36. Alzueta E, de Zambotti M, Javitz H, et al. Tracking sleep, temperature, heart rate, and daily symptoms across the menstrual cycle with the Oura ring in healthy women. Int J Womens Health 2022;14:491-503. [doi: 10.2147/IJWH.S341917] [Medline: 35422659]
37. Chee NIYN, Ghorbani S, Golkashani HA, Leong RLF, Ong JL, Chee MWL. Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. Nat Sci Sleep 2021 Feb 15;13:177-190. [doi: 10.2147/NSS.S286070] [Medline: 33623459]
38. Asgari Mehrabadi M, Azimi I, Sarhaddi F, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 2;8(10):e20465. [doi: 10.2196/20465] [Medline: 33038869]
39. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. Behav Sleep Med 2019;17(2):124-136. [doi: 10.1080/15402002.2017.1300587] [Medline: 28323455]
40. Trasande L, Ghassabian A, Kahn LG, et al. The NYU Children's Health and Environment Study. Eur J Epidemiol 2020 Mar;35(3):305-320. [doi: 10.1007/s10654-020-00623-6] [Medline: 32212050]
41. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108. [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
42. Sivertsen B, Hysing M, Dørheim SK, Eberhard-Gran M. Trajectories of maternal sleep problems before and after childbirth: a longitudinal population-based study. BMC Pregnancy Childbirth 2015 Jun 2;15:129. [doi: 10.1186/s12884-015-0577-1] [Medline: 26031504]
43. Alzueta E, Gombert-Labedens M, Javitz H, et al. Menstrual cycle variations in wearable-detected finger temperature and heart rate, but not in sleep metrics, in young and midlife individuals. J Biol Rhythms 2024 Oct;39(5):395-412. [doi: 10.1177/07487304241265018] [Medline: 39108015]
44. Masaki M, Tsumoto S, Tani A, et al. Discrepancies between subjective and objective sleep assessments revealed by in-home electroencephalography during real-world sleep. Proc Natl Acad Sci U S A 2025 Jan 21;122(3):e2412895121. [doi: 10.1073/pnas.2412895121] [Medline: 39819218]
45. Grandner MA, Malhotra A. Sleep as a vital sign: why medical practitioners need to routinely ask their patients about sleep. Sleep Health 2015 Mar;1(1):11-12. [doi: 10.1016/j.sleh.2014.12.011] [Medline: 27840846]
46. Nielsen LM, Hauge MG, Ersbøll AS, et al. Women's perceptions of cardiovascular risk after preeclampsia: a qualitative study. BMC Pregnancy Childbirth 2022 Nov 11;22(1):832. [doi: 10.1186/s12884-022-05179-9] [Medline: 36368968]

XSL•FO
RenderX

Abbreviations

AHA: American Heart Association
AIM: Acceptability of Intervention Measure
BP: blood pressure
CHES: Children's Health and Environment Study
CVD: cardiovascular disease
CVH: cardiovascular health
HDP: hypertensive disorders of pregnancy
LE8: Life's Essential 8
NYU: New York University
PSQI: Pittsburgh Sleep Quality Index

Edited by A Coristine; submitted 22.Jul.2025; peer-reviewed by A Kozai; revised version received 29.Apr.2026; accepted 29.Apr.2026; published 26.May.2026.

Please cite as:
Hausvater A, Shwayder E, St-Onge MP, Kahn LG, Plazas-Montana M, Na L, Joa A, Trasande L, Reynolds HR
Sleep and Cardiovascular Health Among Women With a History of Hypertensive Disorders of Pregnancy: Pilot Observational Study
JMIR Cardio 2026;10:e81118
URL: https://cardio.jmir.org/2026/1/e81118
doi:10.2196/81118

© Anais Hausvater, Elianna Shwayder, Marie-Pierre St-Onge, Linda G Kahn, Manuela Plazas-Montana, Lillian Na, Amanda Joa, Leonardo Trasande, Harmony R Reynolds. Originally published in JMIR Cardio (https://cardio.jmir.org), 26.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceived Potential and Challenges of Supporting Coronary Artery Disease Treatment Decisions With AI: Qualitative Study

Khara Sauro[1,2,3], PhD; Bishnu Bajgain[3], MSc; Cody van Rassel[4,5], PhD; Bryan Har[5], MD, MPH; Robert Welsh[6], MD; Joon Lee[3,4,5], PhD

[1]Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[2]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[3]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[4]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, 3280 Hospital Dr NW, Calgary, AB, Canada
[5]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[6]Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Joon Lee, PhD
Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Abstract

Background: Coronary revascularization decision-making for patients with coronary artery disease (CAD) can be complex and challenging. Artificial intelligence (AI) has the potential to improve this decision-making by bringing data-driven insights to the point of care.

Objective: We aimed to elicit, collect, and analyze various stakeholders' perceived potential and challenges related to developing, implementing, and adopting AI-based CAD treatment decision support systems.

Methods: A facilitated small-group discussion method, known as a World Café, was conducted with general cardiologists, interventional cardiologists, cardiac surgeons, patients, caregivers, health system administrators, and industry representatives. One-on-one interviews were conducted for participants who could not attend the World Café. Perceived potential and challenges of AI-based CAD treatment decision support systems were solicited by asking participants three broad questions: (1) What is most challenging about revascularization decision-making? (2) How could an AI tool be integrated into the existing clinical workflow? (3) What are the critical components that need to be considered when developing the AI tool? Thematic analysis was performed to identify themes from the data.

Results: Nine participants completed the World Café, and 3 participants completed the one-on-one interviews. Five main themes emerged: (1) evidence-based care, (2) workload and resources, (3) data requirements (subthemes: patient-centered approach, evidence-based AI, and data integration), (4) tool characteristics (subthemes: end user built; generation and presentation of decision support information; user-friendliness and accessibility; and system logic, reasoning, and data privacy), and (5) incorporation into clinical workflow (subthemes: AI as an opportunity to improve care and knowledge translation).

Conclusions: While health care providers aim to provide evidence-based care, CAD treatment decision-making can often be subjective due to the limited applicability of clinical practice guidelines and randomized controlled trial evidence to individual patients. AI-based clinical decision support systems may be an effective solution if the development and implementation focus on the issues identified by end users in this study (patient preference, data privacy, integration with clinical information systems, transparency, and usability).

(JMIR Cardio 2026;10:e81303) doi:10.2196/81303

KEYWORDS

coronary artery disease; clinical decision support; artificial intelligence; technology adoption; implementation science; stakeholder engagement

Introduction

Coronary artery disease (CAD) is characterized by reduced blood flow to the heart muscles caused by plaques in the coronary arteries. The gold standard diagnostic procedure for CAD is coronary angiography performed in cardiac catheterization laboratories, using radiocontrast agents and x-rays to diagnose the disease. Typically, the treatment decision involves determining whether the problematic coronary arteries need to be revascularized via either percutaneous coronary intervention or coronary artery bypass graft (CABG) surgery,

XSL•FO
RenderX

or whether the most appropriate treatment is medical therapy only.

While clinical practice guidelines based on randomized controlled trials exist [1], coronary revascularization decision-making can be complicated by complex CAD (eg, multivessel disease), challenging coronary anatomies, comorbidities, unique patient characteristics, and patient preferences. Although multidisciplinary Heart Team approaches, where diverse specialists, including general cardiologists, interventional cardiologists, and cardiac surgeons, discuss the patient case and formulate the best treatment as a group, are recommended for revascularization decision-making for complex CAD [2,3], they are neither standardized nor evidence-based, making it difficult to operationalize complex treatment decision-making systematically.

Artificial intelligence (AI) has the potential to support coronary revascularization decision-making via data-driven insights. By leveraging patterns and relationships learned from a large amount of patient data, AI models can generate and present personalized decision support insights at the point of care. However, even if AI models with good performance are available for deployment, their technical and operational implementation in real-world clinical environments remains challenging and requires adoption from a variety of stakeholders, including patients, clinicians, health system administrators, health care payors, researchers, and developers [4]. Understanding the barriers and enablers to adopting AI-based clinical decision support systems (CDSS) is critical to developing and implementing such systems in clinical practice.

This study aimed to understand the perceptions of how an AI-based CDSS can facilitate CAD treatment decision-making.

Methods

Study Design and Setting

A World Café [5] was used to elicit and collect stakeholder perceptions about the use of an AI-based CDSS for CAD treatment decision-making. A World Café is a formal, semistructured method that engages diverse stakeholders through multiple rounds of small group discussions, each guided by a targeted question. It is designed to create an open, café-style atmosphere that encourages equitable participation and the flow of ideas across groups. This method enables intimate discussions among participants with varied perspectives, supporting the identification of themes relevant to a topic.

This study intended to complete a single World Café; however, some of our key end users (clinician participants) were unable to attend the World Café due to unexpected urgent patient cases. To ensure that these participant perspectives were captured in the dataset, 3 additional one-on-one interviews were conducted at a later date using the same semistructured protocol and targeted questions used in the World Café.

The World Café and interviews took place in Alberta, Canada, from May 2, 2022, to July 18, 2022, using the videoconferencing and online meeting platform, Zoom (Zoom Communications). Only participants and researchers were present during the meetings and interviews.

Participants and Recruitment

Study participants were recruited in Alberta and included: (1) clinicians involved in the care of patients with CAD including general cardiologists, interventional cardiologists, and cardiac surgeons; (2) health system administrators; (3) private-sector representatives from the cardiovascular information system industry; and (4) patients and caregivers (individuals aged 18 y and older with CAD, or caregivers supporting individuals with CAD).

All clinician participants were practicing physicians within Alberta Health Services (AHS), which is one of the largest fully integrated provincial health systems in North America. AHS oversees centralized delivery of acute care, emergency medical services, diagnostics, and many community-based programs for over 4.4 million Albertans. Although Alberta operates within Canada's universal, publicly funded health care system, its structure differs from many jurisdictions in Canada and abroad by unifying services under a single provincial authority rather than regionally or privately administered systems [6].

Participants were identified using purposive sampling to maximize variation in backgrounds and sex differences. Potential participants (World Café participants [WCPs] and one-on-one interview participants [OIPs]) were recruited using our research network and invited to participate via email. Once potential participants indicated an interest, the consent script was sent to them via email. They were provided with multiple opportunities to ask questions before completing the oral consent process. The consenting process occurred before the World Café and one-on-one interviews.

Data Collection

Consistent with the World Café methodology [7], data were collected through facilitated discussion on broad questions. Three questions were posed to both the WCPs and OIPs sequentially: (1) What is most challenging about revascularization decision-making? (2) How could an AI tool be integrated into the existing clinical workflow? (3) What are the critical components of the AI tool that need to be considered when developing the tool?

A facilitator guided the discussion and used prompts to generate discussion for each question. A note-taker collected field notes to document the context of the discussion (eg, the physical environment and individuals' nonverbal communication) and captured a summary of the discussion, which was shared with the participants at the end of the session (member checking). The World Café session and interviews were digitally recorded and transcribed verbatim, and the field notes were incorporated into the transcripts for analysis.

Reflexivity

The World Café and interviews were conducted by trained, experienced male and female facilitators who had formal graduate-level and experiential training in qualitative methodology and interview facilitation. Several members of the research team, including the facilitators and principal

XSL•FO
RenderX

investigators, had pre-existing professional and nonprofessional relationships with some participants, which may have influenced rapport and data interpretation. One principal investigator was a family member of a patient with CAD who had recently been diagnosed and had an urgent CABG. During the analysis phase, these pre-existing relationships and researchers' personal backgrounds were explicitly discussed during team meetings to reflect on how personal experiences, disciplinary backgrounds, and expectations may shape data interpretation and the construction of final themes.

Data Analysis

Transcripts from both the World Café and one-on-one interviews were uploaded, managed, and analyzed using NVivo (version 12.0, Lumivero). The Clarke and Braun [8] approach to thematic analysis was used to analyze the data. An inductive approach was used to identify codes and themes from the data. A first analyst familiarized themselves with the data and identified and established codes in a coding book as a reference. A second independent analyst then familiarized themselves with the data and reviewed the preliminary codes identified by the first analyst, revising and adding new codes while interpreting the data as a circular process that moved back and forth between smaller parts of the transcript and the whole text. This iterative coding process was applied as new themes emerged, and the transcripts were reread to verify that the codes and themes were not missed. These coding "nodes" were discussed among the research team and then consolidated into themes [8]. Coding discrepancies between analysts were addressed through discussion and joint review of the relevant transcript segments. Consensus was achieved through iterative comparison of interpretations, and any disagreements were resolved collaboratively to ensure consistent application of codes across the dataset.

In contrast to the group dynamic characteristic of the World Café, individual interviews may elicit more detailed and individualized reflections. To reconcile these methodological differences and enable direct comparison between formats, all transcripts were coded using the same unified coding framework. Codes and themes were examined for convergence and divergence, and only themes supported by patterns across both data sources were used in the final analysis. Although the one-on-one interviews were conducted following the World Café, the interview guide was not refined or modified based on World Café findings, and the same semistructured protocol was used across formats.

Trustworthiness

Various strategies were used to ensure the trustworthiness of the findings [9]. The transcripts were reviewed by the World Café and interview facilitators for accuracy before analysis. We used member checking at the end of the World Café and each interview by summarizing the discussion and asking participants if we accurately captured the discussion. Regular peer debriefing and discussion took place between members of the research team about the representation of this study's population, recruitment, data collection strategies, and data analysis, from the data coding process to the emerging themes, to enhance the accuracy of the results. The results were reviewed and refined by all authors, some of whom were participants.

Ethical Considerations

The University of Calgary Conjoint Health Research Ethics Board approved this study (REB20-1879). Before participation, explicit oral informed consent was sought and obtained from all study participants. The privacy and confidentiality of participants' data and identity were maintained by following the approved research data security and privacy protocol. Participants were not compensated. Additionally, this paper follows the COREQ (Consolidated Criteria for Reporting Qualitative Research) guideline (Checklist 1) [10].

Results

Overview

The World Café was conducted on May 2, 2022, and the interviews were conducted between June 15 and July 18, 2022. The World Café lasted about 120 minutes, and each interview ranged from 45 to 60 minutes. The participants (9 male; 3 female) were cardiologists (n=4), interventional cardiologists (n=2), cardiac surgeons (n=2), health system administrators (n=1), patients and caregivers (n=2), and an industry representative (n=1). Clinician participants (n=8) included early-career (n=3), middle-career (n=4), and senior professionals (n=1).

Five overarching themes emerged from the data: (1) evidence-based care, (2) workload and resources, (3) data requirements (subthemes: patient-centered approach, evidence-based AI, and data integration), (4) tool characteristics (subthemes: end user built; generation and presentation of decision support information; user-friendliness and accessibility; and system logic, reasoning, and data privacy), and (5) AI incorporation into clinical workflow (subthemes: AI as an opportunity to improve care and knowledge translation). Each theme is described in detail in the ensuing sections, with example quotes tabulated in Tables 1-5 for each theme. Quotations are identified by stakeholder group and by data collection source (ie, WCPs or OIPs).

XSL•FO
RenderX

Table . Example quotes related to evidence-based care (theme 1).

Description	Example quotes
The importance of evidence-based clinical practice.	• "We need more evidence when the areas we do have evidence still aren't standardized… Could more evidence help standardize things? And I do want to say that there are areas, certainly, that we need more evidence, but just like the counterbalance, there is, there are areas where we have evidence, and it still hasn't standardized practice." [OIP #1, surgeon] • "It certainly does. And I think it would for everybody. Everybody should be thinking of the guidelines, but it's kind of a starting point because it's often the nuances, or there's the other clinical variables that aren't in the classic guidelines that are important considerations." [OIP #3, interventionalist] • "It can be challenging identifying which patients you're going to want to proceed with revascularization versus proceeding with medical therapy… it does become somewhat knee-jerk that a person has in anginal symptoms or they have a non-invasive test that's suggestive of ischemia, and automatically they get sent to the cath lab with the thought that they're going to be revascularized. Now, the existing literature and the existing guidelines don't actually support that… it can be challenging…" [OIP #2, interventionalist]
Uniform and standardized decision-making between clinicians: different priorities based on their values & success rates (PCI[a] vs CABG[b]).	• "I think most decisions for revascularization are made ad hoc, on the spot, and I think that's reasonable for most of the time, but it really is the setup for a practitioner dependent practice… I think all of us see variation in practice, and I think all of us see that there's areas that aren't standardized… It would be nice if things were standardized, it simplifies things and allows everyone to ensure that we're aligned or at least the expectations are clear." [OIP #1, surgeon] • "I think there is variation in practice. But, say, between different surgeons, there's definitely variability. But, ultimately, say, if it's in the middle of the night, then it's basically what my preference or opinion is, I guess, at that point. What my colleagues might do might be different, but it doesn't really affect my decision process at that point." [OIP #3, interventionalist]

[a]PCI: percutaneous coronary intervention.

[b]CABG: coronary artery bypass grafting.

XSL•FO
RenderX

Table . Example quotes related to workload and resources (theme 2).

Description	Example quotes
The impact of decision-making process due to workload and physician burnout. Lack of time to explore and discuss patient information due to high volume of work.	• "When I have been burnt out, of course that impacts my practice, of course that impacts my decision making. I imagine that each of these... I think there's a very high resiliency rate within each of these groups, and I think, a lot of self-awareness to monitor burnout. It's impossible for me to quantify the impact, I'd just be speculative, but I think that all of us have to be mindful of that impact in decision making." [OIP #1, surgeon] • "When we're burnt out and we're overloaded sometimes, there's a tendency towards the path that's going to give us a more definitive answer more quickly." [OIP #2, interventionalist] • "The other considerations are the timing of revascularization. Is there active ischemia at the time? Is it an emergency that needs to be done right away? Or is it something that can wait until something else is optimized, either antiplatelet strategies or anticoagulation or other patient variables? Clinical status? Timing on like ... Is it emergency or not? Or is it urgent? Or is it elective?" [OIP #3, interventionalist] • "I was thinking about time as being one of the biggest issues for me and also acuity. So, I find that you have more time to weigh those risks and benefits and do your own cost-benefit analysis in the more stable scenario. But, in the acute phase, that time to do that goes at the side of trying to get the intervention going and the case started… It should be standardized for all patients, regardless of the presentation. But in my own individual experience, I find that time is really an issue, and it depends on the urgency of the scenario." [WCP #6, cardiologist] • "One is synthesizing all of the data that comes in. So, whether it's the anatomy, the patient comorbidities presentation, and what everyone's perspective on feasibility of getting a good result, whether it's by angioplasty or bypass surgery. So, I think it's putting everything together… you're trying to get through all of these patients, being comfortable that you've gathered all the relevant data to make the right decision is quite difficult…" [WCP #5, interventionalist] • "We might get a mailing list the day before…, you might spend a little bit of time in the evening reviewing the angiogram films. And the next day they might only spend five minutes on a patient… So, you're trying to get through all of these patients, being comfortable that you've gathered all the relevant data to make the right decision is quite difficult, given the fact that a lot of times you have never met the patient… So, certainly that you can get time pressure and not discuss patients thoroughly enough." [WCP #5, interventionalist]

XSL•FO
RenderX

Description	Example quotes
Considering cost of care (PCI[a] vs CABG[b]) vs value of care.	• "Back to our patient's perspective, I think we do look at risks of the non-discussed pieces of it when we make our decisions with patients. So, obviously as clinicians, we talk about death or MI stroke, but there are a lot of other things that go into the decision-making like time in hospital, recovery status… we would want to discuss with the patient to make sure they understand. So, at 10 years you may have a slightly better risk of death, but in the meantime, you've got a recovery period that would be hard to manage. So, cost may be less so than the patient discussions, but as the healthcare system tightens, cost will become more and more important in the future." [WCP #4, interventionalist] • "We do talk about value, defining value as outcome over cost. So, it's not that the cost doesn't matter. And I think the system is willing to pay the cost as long as the outcome achieved from that cost makes sense. So, I think in our system, we would talk more about value" [WCP #3, administrator] • "When you're pressed for beds if the results are close enough between bypass and PCI, does your resource limitation push you more towards one or the other?... And so, certainly we did have those discussion... you just deal with one or the other either with surgery or angioplasty, then the fact that five years down the road they might be back for a second procedure. But you just try to keep the resources freed up in the system, whether it's time or money or beds." [WCP #5, interventionalist] • "Everything that we're talking about is incredibly expensive… I would think, in the scheme of budget for this, this would be relatively small, and anything that makes things more efficient probably will save money more than whatever it costs." [OIP #3, interventionalist]
Considering resources (e.g., cost, investment), and willingness to invest and implement the AI[c] tool.	• "If we want these tools to be adopted, they have to be purchased… a lot of the times we are making those arguments as a return on investment, essentially saying that if they pay this much for this software, it will save them time, it will lead to improvements in quality for value-based reimbursement or these sorts of things… So, I do think there's the clinical perspective in terms of, directly with the patient and beds and everything. And then there's also the administrative perspective around time and resources and whether they are willing to put in the effort to implement this in practice in order to actually see the results. I think it's an important consideration, even this early on in the process." [WCP #9, industry partner]

[a]PCI: percutaneous coronary intervention.

[b]CABG: coronary artery bypass grafting.

[c]AI: artificial intelligence.

XSL•FO
RenderX

Table . Example quotes related to data requirements (theme 3).

Subtheme and description	Example quotes
Patient-centered approach	
Complex patients with multimorbidity	• "I think one of the bigger challenges that's happening more and more now is our patients are older. They're more complex. They have more comorbidities. The risks of everything are higher. Their disease is getting more complex. There are sometimes not reasonable PCI options… And so, you are not infrequently trying to treat a patient who has complex multi-vessel disease, who is a poor candidate for surgical revascularization, or a very high-risk candidate for surgical revascularization, who is equally a high-risk candidate for percutaneous revascularization." [OIP #2, interventionalist] • "There are the anatomic things like the coronary anatomy. There are the other comorbidities are a big role... I guess the factors that relate to their potential benefits. Are they potentially receiving a symptom or a survival benefit? And then what are their risk factors?... Because it's always a balance of benefit versus risk." [OIP #3, interventionalist]
Understanding patient expectations and preferences	• "If you give the patients a choice, they'll all pick PCI. And PCI is great for a lot of them, but it's not right for everybody. And that's where that education has to come in, and then they can make their decision." [OIP #3, interventionalist] • "I chose a course of treatment that had the lowest chance of incontinence over the others. So, it's weird, I'm not sure patients really think exactly like clinicians in this situation. I was opting for quality of life over other factors, and it served me well… But those would be important to me." [WCP #2, patient] • "Nobody wants to have surgery, but people are interested in their long-term outcomes. They don't want repeated heart attacks… the second one is also understanding what the patient's preference is and how much of that weighs on the decision. So, if it's 60% in favor of bypass and 40% in favor of angioplasty from a clinician standpoint, but what if the patient feels very strongly that they want to have angioplasty and you would need to tell them various substantial risk… or substantial benefit of bypass surgery. So, understanding what the patient preference is based upon what would happen if they were presented with the data." [OIP #3, interventionalist]
Evidence-based AI[a]	
Representation of diverse patients in the data used to develop the model	• "I think that making sure that you want the largest pool of data possible, but making sure that data represents a wide cross-section of demographics, and that you're not trying to ultimately end up generalizing it to populations that haven't been involved in generating the model. That's really important." [OIP #2, interventionalist]

XSL•FO
RenderX

Subtheme and description	Example quotes
Complementing and improving upon the existing clinical practice and evidence	• "I would love to see both strengths of AI put to work… there are a lot of things we know, and honestly, I think we're maybe not that bad at this, but it would be great to see if AI comes up with information or predictors that we didn't know, which would be the real beauty of this model." [WCP #4, interventionalist] • "You're going to do the standard stuff that we all think about, age, diabetes, hypertension, dyslipidemia, frailty, down the list, but the black box approach where you basically ask the computer to tell us what's in the model, I think would be very interesting and potentially the validation moving forward would be very exciting." [WCP #4, interventionalist] • "AI-based technologies is to be able to give us information that some standard studies don't give us… It may be possible that by applying more complex machine learning, you might actually be able to find an answer that we haven't been able to find so far in our existing evidence. So, I think that you have to be aware of what the existing evidence is, but if you find a different answer using new technology, well, I don't think you adopt it wholeheartedly and ignore what we have figured out already, but I think you need to integrate that in." [IP #2, interventionalist] • "The evidence is coming typically from trials that have a lot of nuances to them. Applying guidelines or a specific trial to a specific patient can be real challenge because they don't always directly apply. That's where the clinical judgment and oversight and gestalt, I think, play a role and is trying to say, well, the evidence that we have ... How well does it apply to this patient? And that's where they don't always directly apply." [IP #3, interventionalist]
Data integration	

XSL•FO
RenderX

Subtheme and description	Example quotes
Ethics, privacy, and confidentiality	• "I think that from a patient privacy perspective, that's going to be a really important thing to have a really good grasp on the ethics before that happens, because machine learning can be used to generate all sorts of models for all sorts of risks. And whether those things in the future are going to... Say, my EMR is plugged into machine learning, and it's able to automatically generate a detailed risk for, for cardiac death over the next 10 years, is that going to affect my patient's ability to get insurance? So, there needs to be a consent aspect in there for sure." [OIP #2, interventionalist] • "Obviously, confidentiality and privacy are massive when it comes to medical information. It certainly would have to be secure." [OIP #3, interventionalist]
Integration with the EMR[b].	• "Something that can be incorporated directly into the EMR, so you can just say, 'Oh, well, I want to calculate the whatever score for this patient,' so you can click on it and it can pull in whatever data that it needs. I think that the simpler you make it, the more likely you're going to see uptake." [OIP #2, interventionalist] • "When we talk about how this could be integrated in the existing workflow, I think we have to consider Connect Care as part of this…, Healthcare institutions have put significant investment into the implementation of electronic medical record systems… which opens up an opportunity for us, if we can integrate with those systems and make it fairly seamless experience." [WCP #9, industry partner]
Integration of comprehensive data elements	• "The critical components one is, as everyone's mentioned, is the APPROACH CARAT diagram and all the different components of it, whether it's the anatomy or the jeopardy score, lesion characteristics, previous stents, the comorbidities that are the classic comorbidities, whether it's renal failure or diabetes, left ventricular dysfunction, patient age, BMI, whether it's super high or super low. The ones that are harder to capture, I think, beforehand… other one is frailty… it's something that's really hard to capture, but I think the frailty piece is really important." [WCP #5, interventionalist]

[a]AI: artificial intelligence.

[b]EMR: electronic medical record.

XSL•FO
RenderX

Table . Example quotes related to tool characteristics (theme 4).

Subtheme and description	Example quotes
End user built	
Engaging end users throughout the development and implementation processes	• "When developing something that is intended to be used as a tool for someone, that person is the stakeholder and they should be engaged and heavily involved in all aspects of the development, training, implementation, and subsequent follow up and iterations. It's hard for me to imagine a step a clinician shouldn't be involved…" [OIP #1, surgeon] • "Physicians and patients would be the end-users for those models that are generated. So, matter of asking what more information the end-users actually need." [OIP #2, interventionalist] • "I think something like this could be complicated enough that, without a physician or surgeon or an interventionist… clinical feedback would be important in order to tailor it to what the clinicians need to see." [OIP #3, interventionalist] • "What data are available, what data are easy to get, which ones would require the physician to do extra steps? And maybe prioritizing ones that are accessible or prevalent versus ones that might require extra work and therefore prevent adoption. So, I think that's, figuring out what we can do with Connect Care or very similar systems is important here." [WCP #9, industry partner]
Engaging patients throughout and capturing patients' voices	• "If the AI tool was developed with the patient-centered focus with patient researchers actually involved in developing the tool. And it might provide a communication roadmap for some of your colleagues… and coach them in both the language and what the patient is actually looking from their clinician to provide in the way of information, so that the discussion is actually much more robust." [WCP #4, interventionalist] • "Asking the patients how they're doing, their mobility, quality of life, frailty, we know from other research that's been done that PROMs are quite predictive of outcomes, and they come out, when you run the AI with all the different features, PROMs often come out as contributing to that final prediction." [WCP #9, industry partner]
Family physicians as potential end users	• "Is that possible that it could be starting with a family physician instead of specialist cardiologist? I don't know where else it potentially could start. That way, it's an easier conversation for the family physician to have with the patient." [WCP #1, patient] • "I would fully agree with that (integrating AI in PHC)… The only thing is that some family physicians are overwhelmed by the breadth of knowledge they need to know about your diabetes, your heart's arteries, your medical therapy, and they really look to the cardiologists and the surgeons, where appropriate… the intricacies of decision-making." [WCP #4, interventionalist]
Generation and presentation of decision support information	

XSL•FO
RenderX

Subtheme and description	Example quotes
Need for scores as a way to summarize granular information for easy interpretability and communication	• "Making decisions about revascularization would be the development of more complex risk-prediction models… through machine learning, there would be the ability to generate a more sophisticated model for helping to estimate risk. And again, you're not going to ultimately have an AI making the decision about whether or not you're going to proceed with revascularization, but to be able to go into an interaction with a patient and be able to give them more granular information about what their risks are, that could be helpful." [OIP #2, interventionalist] • "I think it'll be determining what all risk factors need to go into this tool (AI) and what weight are you going to give to each risk factor and does presentation…, How do you give mobility a number, and comorbidities… but some of these risk factors are going to be a little bit difficult to try to quantify." [WCP #7, cardiologist] • "My first thought was on the APPROACH diagram, there's a jeopardy score and the jeopardy score is used, to a degree, to guide our revascularization decisions. It would almost make sense that this prediction tool puts some sort of either score or recommendation was my first thought." [WCP #4, interventionalist] • "Overlay how AI has actually helped populate the data in the various columns of chart data. That, to me, I think would be perhaps a good marriage of data… on the various success factors that you're striving towards by each treatment option. And then that would provide a good discussion basis for the physician and the patient and their family… if you stack them all together, you would have what would logically be the optimal treatment plan, just because the higher score and whatever it would add up, it would be ably demonstrated… it's easy to explain at the bedside, I think would be a big benefit rather than talking in clinical jargon." [WCP #2, patient]
Importance of data quantity and quality	• "I think that's how you would harness the true power of AI and machine learning. As much data as you could get in, I think that's how you could really…harness the power of this method." [OIP #1, surgeon] • "In order to get high quality information out of any machine learning system, it's entirely based on the volume of data you're able to provide it… I think the broader the data you're able to plug into a machine learning system, the better, because one of the big issues that I'm aware of with machine learning is just the impact of the bias of the information that you put in." [OIP #2, interventionalist] • "As a surgical resident who's doing a lot of research with clinical data, being able to extract that data efficiently would be key... I think we've struggled a little bit with the comprehensiveness of data and the efficiency of it." [WCP #8, surgeon]
User-friendliness and accessibility	

XSL•FO
RenderX

Subtheme and description	Example quotes
Importance of an easy-to-use, intuitive user interface and automation.	• "User interface are really important… I'm not still clear how to evaluate an algorithm… I don't have a good sense as to how to actually perform due diligence on a product or algorithm. And in the absence of the ability to do that, it's very difficult to know how much weight you would put on the response from an algorithm. So, I think that a bit of knowledge translation, or trying to validate an AI model in a manner that's understandable to clinicians is critical." [OIP #1, surgeon] • "I think a lot of what it comes down to is really just convenience, and speed, and ease of use… certainly, automated measurements, automated tracing of the left ventricle and calculation of, for example, a 3D left ventricular ejection fraction. These are areas that accuracy is improved… so I think usability becomes really, really important, particularly at a time when technology has in certain areas of medicine, led to a lot of added complexity with an unclear value proposition. And so I think that really providing an easy to use interface and easily digestible material is our answers, is really helpful." [OIP #1, surgeon] • "When I think about kind of AI tools, they're going to be things that physicians reach for when they want them… you're going to be reaching for every patient… for something as simple as like MDCalc to put in someone's Framingham risk score. So, you would want something that is easily accessible, something that is smartphone-based, or even something that can be incorporated directly into the EMR, so you can just say, 'Oh, well, I want to calculate the whatever score for this patient,' so you can click on it and it can pull in whatever data that it needs. I think that the simpler you make it, the more likely you're going to see uptake" [OIP #2, interventionalist] • "In general, surgeons aren't very technically or at least computer savvy. The simpler, the better, for sure. Because not too many surgeons will commit to a lot of learning for something like that, computer work, or have a lot of time to commit to it. Actually, it is quite important that it's something that's intuitive or easy to work with. Being user friendly, for sure, is important. Being accessible. Something that we can use remotely is important because that's when, I think, it'd be pretty useful. And then just current, I guess. Data or information that's real time." [OIP #3, interventionalist]
System logic, reasoning, and data privacy	

XSL•FO
RenderX

Subtheme and description	Example quotes
Transparency and accountability	• "I think that there does need to be some sort of transparency in how the tool is working… For decisions, an overall gestalt is used a lot... is this person a candidate for this? Is this person ... Would it be ... Would they do well with this? And so, the gestalt is an important part. With artificial intelligence, I don't know if gestalt is a part of it. That would be a challenge, I would think. Having some transparency, though, in how the algorithms or whatever it is that are making recommendations or decisions or things like that ... A transparency so that people can see how this tool comes to these conclusions." [OIP #3, interventionalist] • "I think the only thing that we have to be fully aware of and cognizant of… from a medical legal point of view, having a statement that X is the desired outcome without all of the qualifying pieces, if the Y procedure is done and the patient has a bad outcome, which may or may not be predictable, usually not predictable, it may lead to a medical legal nightmare… So, we just have to be cognizant of that and decide where this decision tool lands in terms of its availability and to who." [WCP #4, interventionalist]
Living documents and real-time feedback	• "From a workflow perspective, AI really needs to be almost synchronous with the test itself and providing real time feedback or near time feedback to be really clinically useful." [OIP #1, surgeon] • "I think it needs to be part of the living document and available at the time we do the angiogram to be most valuable. I would say that 90% of the time the revascularized option, which is anatomically best for the patient and comorbidly best for the patient, taking their patient profile is obvious 90% of the time. It's either a straightforward stent, we deal with it, move on. It's either ongoing medical therapy, we optimize that and move on, or it's a clear-cut patient that should be moved forward for bypass..." [WCP #4, interventionalist] • "Being user friendly, for sure, is important. Being accessible. Something that we can use remotely is important because that's when, I think, it'd be pretty useful. And then just current, I guess. Data or information that's real time." [OIP #3, interventionalist]
Data privacy and confidentiality	• "I think that from a patient privacy perspective, that's going to be a really important thing to have a really good grasp on the ethics before that happens, because machine learning can be used to generate all sorts of models for all sorts of risks. And whether those things in the future are going to... Say, my EMR is plugged into machine learning, and it's able to automatically generate a detailed risk for, for cardiac death over the next 10 years, is that going to affect my patient's ability to get insurance? So, there needs to be a consent aspect in there for sure." [OIP #2, interventionalist] • "Obviously, confidentiality and privacy are massive when it comes to medical information. It certainly would have to be secure." [OIP #3, interventionalist]

XSL•FO
RenderX

Table . Example quotes related to AI[a] incorporation into clinical workflow (theme 5).

Subtheme and description	Example quotes
AI as an opportunity to improve care	
Considering AI as an opportunity in the existing clinical workflow.	• "For sure, there's many opportunities… I think risk stratifying the lesion or identifying high risk lesion characteristics… identifying a culprit lesion, the predicted success of revascularization or subsequent stent complications. In another way of saying that would be determining what method of revascularization. I think those are all-important real-time feedback that an operator could receive… It would be really nice to pair that with non-invasive cardiac diagnostics, which would include echo, MRIs… Those are all areas that I think will be very, very interesting." [OIP #1, surgeon] • "I think that having some artificial intelligence to make suggestions is helpful. And like I said, to provide supporting information like risk profiles and stuff like that, but I think that ultimately, the conversation between the physician and the patient is always going to be a pretty big driving force for what path we go down." [OIP #2, interventionalist]
Knowledge translation	
The importance of knowledge translation among end users.	• "One of the challenges for us in knowledge translation here, trying to translate something that's extremely technical. And in fact, we're front-runners in the industry into something that's useful for both patients and caregivers. So, I don't think it's a game stopper, it's just something we have to be aware of and be knowledgeable about." [WCP #2, patient] • "You guys do such important work, really the patient should be knowing this stuff, like what's going on with their body at a doctor's visit, at a family physician. So, I think if this information's being shared." [WCP #1, patient] • "I don't have a good sense as to how to actually perform due diligence on a product or algorithm. And in the absence of the ability to do that, it's very difficult to know how much weight you would put on the response from an algorithm. So, I think that a bit of knowledge translation, or trying to validate an AI model in a manner that's understandable to clinicians is critical." [OIP #1, surgeon]

[a]AI: artificial intelligence.

Theme 1: Evidence-Based Care

Overview

All participants emphasized the importance of evidence-based guidelines to minimize variation in care and unify the Heart Team that makes treatment decisions for patients with CAD. The clinician participants explained that their clinical practice was based on their clinical knowledge and experiences, as well as existing medical evidence (including clinical practice guidelines). They stated that although there are some clinical guidelines, many are of low quality and endorsed the need for better guidelines. They also noted that while evidence and experience are foundational, CAD treatment decisions tend to be biased by the opinion of a single clinician and that each clinician has different priorities and experiences.

The clinician and health system administrator participants noted that, in addition to the evidence informing treatment decisions, factors such as the urgency of the case, provider workload, patient preferences, and time of day also influence treatment decisions. This led to comments related to timely and complete patient data (medical history and comorbidities), which is discussed in greater detail under theme 3 (data requirements).

Contrasting Perspectives

While all participants endorsed the importance of evidence-based care, clinicians primarily framed the issue as a challenge of guideline clarity and individual bias. In contrast, it was suggested that administrators may instead focus on the broader system-level influences that contribute to variability in revascularization decisions (explored in further detail in theme 2).

Theme 2: Workload and Resources

Overview

Most clinician participants stressed that time is a significant issue for making treatment decisions around revascularization. They stated that due to the high volume of cases, there is minimal time available to comprehensively review patient information and discuss the patient. This challenge, combined with the urgency required to revascularize patients with CAD, makes it difficult to fully assess all the risks and benefits of each treatment approach and consider patient preferences. All participants endorsed the challenges related to the current strain on the health care system and clinicians.

XSL•FO
RenderX

There was also discussion about physician burnout due to persistently high workloads. Clinician participants expressed that when they are overloaded and burnt out, there is a chance they will make a treatment decision more quickly, without fully considering all factors.

Participants suggested that having access to all relevant information for making treatment decisions in one, easily accessible spot would facilitate decision-making, which might help to improve patient outcomes and resource use. This is further discussed in theme 5 (AI incorporation into clinical workflow). Further noted by the patient participants was the importance of also considering patient preferences, which is discussed under themes 3 - 5 below.

Furthermore, there was some discussion about the cost of care vs the value of care (outcomes). Most of the clinician participants emphasized that they valued results over cost when making decisions about revascularization. However, they did suggest that long-term resource consumption should be considered before making treatment decisions, provided the evidence for effectiveness is comparable (eg, CABG vs percutaneous coronary intervention). The health care system administrator also discussed the issue of the value of care, especially in the current context of a resource-strained health care system.

Considerations of the cost and resources of implementing a new AI system into clinical workflow were discussed. It was noted that adopting AI in the clinical workflow is expensive and needs investment (time and money). However, participants suggested that the investment in AI may be relatively small if it eventually saves time and improves the quality of care. The industry participant expressed the importance of engaging and understanding health system administrators' willingness to invest (time and resources) and implement the AI technology into practice from the early stages of the development process.

Contrasting Perspectives

Although all participants acknowledged the strain created by limited time and resources, clinicians emphasized how high workload, case urgency, and burnout directly affect their ability to thoroughly review patient information and weigh treatment options. Patients highlighted the importance of ensuring that their preferences are considered despite time constraints, whereas administrative concerns remained focused on the value of care despite resource limitations. From an industry perspective, attention was drawn to the investment required to implement AI tools and the willingness of administrators to support such adoption.

Theme 3: Data Requirements

Patient-Centered Approach

All participants agreed that patient characteristics were the most critical factor in making a revascularization decision. The clinician participants indicated that making decisions around revascularization is particularly difficult for complex patients with multimorbidity.

Understanding and considering patient preference was another key factor in making revascularization decisions. The clinician participants indicated that giving patients a choice or following their preferences can be challenging, because the patient's preference can sometimes be highly divergent from the evidence-based recommendation.

All participants agreed that respecting patient preference is important, and a conversation between the clinician and the patient, including hearing the patient's perspective while educating patients about the risks and benefits of each intervention, should guide the decision-making process. It was noted that an AI-based CDSS could facilitate this discussion.

Evidence-Based AI

Participants stated that AI-based tools are expected to be based on evidence. The clinician participants emphasized that integrating existing evidence is important. They were intrigued by data standardization that could help AI to address real and predictable risks efficiently and consistently.

In addition, the clinician participants wondered if AI-based recommendations would differ from the current evidence and how they would reconcile such discrepancies. Despite this concern, they also expressed their excitement toward practicing data-driven revascularization decision-making.

Data Integration

Participants voiced the importance of integrating patient data into AI-based CDSSs. Participants stated that AI tools integrated with the electronic medical record could facilitate the clinical use of large volumes of patient data more efficiently and precisely. All participant groups also stated that integrating all critical patient data, including a comprehensive list of risk factors (eg, patient history, comorbidities, anatomical presentation, and frailty), is essential for AI-based CDSSs. Given the size and comprehensiveness of the data involved, some participants raised concerns regarding ethics, privacy, and confidentiality.

Contrasting Perspectives

All participants emphasized both the importance and the challenges of incorporating patient perspectives into revascularization decision-making. Clinicians suggested that an integrated AI-based CDSS may facilitate clinician-patient discussions and enhance decision-making; however, they also raised concerns regarding patient confidentiality and the potential unintended consequences of risk profiling that may adversely affect patients.

Theme 4: Tool Characteristics

End User Built

All participants emphasized the importance of meaningfully involving all key end users when developing the tool. All participant groups stressed that end users should be involved not only in tool development but also in training, implementation, and evaluation. Furthermore, the patient and clinician participants noted that a patient-centered approach that captures patient voices and their perspectives about clinicians' use of AI-based tools would lead to effective communication between patients and care providers. Some participants identified family physicians as potential end users as well.

XSL•FO
RenderX

Generation and Presentation of Decision Support Information

Participants were interested in the possibility that AI-based CDSSs have the capacity to provide a comprehensive score or recommendation that takes into account patient characteristics to optimize treatment plans and have the ability to interact with patients and families. Summary risk scores were preferred for easy interpretability and communication. Conversely, participants expressed concerns about the validity of the CDSS and including a comprehensive list of potential risk factors. Comprehensive data that integrated patient medical history and comorbidities were perceived as core components for successful AI-based CDSSs.

User-Friendliness and Accessibility

A user-friendly interface and being accessible beyond a networked computer (eg, mobile access) would increase uptake. In addition, participants stated that the CDSS needs to be intuitive and integrated seamlessly within the existing clinical workflow. Technical support would improve the usability and implementation of a CDSS, which would also address the concern about resources and workload associated with the tool (discussed in theme 2). The clinician participants also stressed that revascularization decisions require a high-level summary of accurate information, which must be easy to navigate and access.

System Logic, Reasoning, and Data Privacy

Most of the clinician and patient participants focused on the importance of the transparency of the CDSS and knowing how the CDSS works and generates treatment recommendations, including its logic and limitations.

Some participants also expressed their concern about regulatory compliance, liability, and accountability of the CDSS before implementation can be considered. This concern was based on medical ethics and legal perspectives. However, some participants also felt that if data privacy standards could be met, there was excitement about leveraging the advantages of AI.

Similarly, another major concern expressed by many participants was data privacy, particularly whether the AI-based CDSS could maintain the required confidentiality and privacy of health information. They pointed out that CDSS developers should be aware of various regulatory requirements that protect health information privacy. The patient participants were worried that AI recommendations could have unintended consequences on other health-related issues.

Contrasting Perspectives

Although all stakeholder groups agreed that AI-based CDSSs should be user-friendly, accessible, and developed with meaningful end user involvement, clinicians emphasized medico-legal considerations and the desire for such tools to support real-time decision-making and seamlessly integrate into clinical workflow. In contrast, patients may be more likely to view such tools as a means to facilitate discussion with care providers rather than solely functioning as a real-time decision aid. Both clinicians and patients emphasized the importance of validity and transparency of an AI-based CDSS; however, similar to theme 3, concerns remain regarding patient data privacy.

Theme 5: AI Incorporation Into Clinical Workflow

AI as an Opportunity to Improve Care

Most participants perceived AI as an opportunity in the clinical workflow. They were quite positive and supportive about the development of AI-based CDSSs for revascularization decision-making. The clinician participants felt that AI would add support to their decision-making process, provided that recent scientific evidence is incorporated into the CDSS. Many participants mentioned that integration between the electronic medical record and AI, and end user engagement with both clinicians and patients, would be crucial to integration into the clinical workflow.

Knowledge Translation

Participants underlined knowledge translation as one of the core components required before integrating AI into the clinical workflow. For instance, the clinician participants were intrigued by how AI algorithms work, how comprehensive AI-based risk scores can be, as well as the system logic, reasoning, benefits, and limitations of the technology. However, many participants expressed that they still lacked knowledge about how AI works and suggested that continuous knowledge translation would be helpful.

Participants also noted that identifying end users is critical before integrating AI into the clinical workflow. For instance, a patient participant questioned whether AI-based CDSSs could be integrated into not only cardiac care but also primary care. The participant further expressed that integrating AI into primary care would facilitate conversations between patients and their primary care providers.

Contrasting Perspectives

While some clinicians emphasized the potential role of AI-based CDSSs as real-time supports within existing clinical workflows, others highlighted their value in supporting clinician-patient conversations across care settings. Differences in perspectives were most apparent in relation to knowledge translation, with clinicians emphasizing their desire for a deeper understanding of AI system functionality and validity, while other participants highlighted the importance of making AI-derived information simple and accessible to patients and primary care providers.

Discussion

Principal Results

This study provides an exploratory examination of AHS stakeholder perspectives on the use of an AI-based CDSS for CAD treatment strategy. Across stakeholder groups, participants emphasized that delivering evidence-based care when making coronary revascularization decisions for patients with CAD is often challenging due to conflicting or inadequate clinical practice guidelines. Although AI has the potential to improve the current state of CAD treatment decision-making, the need for timely access to comprehensive patient data through integration with hospital information systems, clinicians' heavy

XSL•FO
RenderX

workloads, data privacy concerns, and end users' desire for transparency regarding how the AI generated a particular recommendation may complicate the successful development, deployment, and adoption of AI-based CDSSs. The likelihood of successful adoption can be enhanced by incorporating scientific evidence and patient preference into decision support information, involving end users (both patients and clinicians) in the entirety of technology development, designing systems that are intuitive and easy to use, and presenting information in a succinct and easy-to-understand manner. In addition, this study's participants also raised the importance of systemic issues, including regulatory requirements, the need for resource commitments from health systems, and health care cost considerations.

Despite these challenges, the participants expressed excitement about AI's potential to improve CAD care. Many acknowledged that they had limited AI knowledge and wanted to be educated in an ongoing manner. They felt that although substantial investments may be required to develop, implement, and adopt AI-based CDSSs, the potential cost savings and improved patient outcomes will likely make it a worthwhile endeavor.

Contrasting Perspectives

Across themes, consistent patterns emerged in how different stakeholder groups framed the challenges and opportunities associated with AI-supported revascularization decision-making. Clinicians primarily focused on the practical realities of care delivery, emphasizing limitations in guideline clarity, time pressure, workload, and difficulties with synthesizing large amounts of data. Many described how an AI-based CDSS could help synthesize complex data, enhance clinical efficiency, and support real-time decision-making. Clinician concerns centered on liability and accountability of implementing AI-based tools, how to reconcile AI recommendations with clinical judgment, and emphasized the importance of patient data security and privacy. While clinicians often focus on mitigating long-term mortality and major adverse event risk, patients emphasized the importance of considering both short and long-term risks and benefits, valuing AI-based CDSS as a tool to help facilitate communication between clinicians and care providers. Industry perspectives highlighted the potential challenges surrounding an AI-based CDSS implementation, emphasizing the need for early alignment with administrative priorities, investment considerations, and demonstrable value to support adoption. Health system administrative perspectives, in contrast, tended to frame decision-making through a system-level lens, prioritizing standardization and value of care, viewing an AI-based CDSS as a short-term cost with potential long-term benefits and cost savings. Finally, differences were evident in expectations around knowledge translation, with clinicians seeking a deeper understanding of transparency, validation, and functionality of AI-based systems, while other participants emphasized the importance of making AI-derived information accessible to patients and primary care providers. Together, these contrasting perspectives highlight that the successful implementation of AI-based CDSSs requires not only technical accuracy but also careful alignment with the distinct priorities, responsibilities, and expectations of diverse stakeholder groups.

Comparison With Prior Work

The potential of AI in improving CAD or cardiovascular care at large has previously been discussed [11,12], and a number of machine learning models have been developed in this space for a variety of clinical use cases (eg, in other studies [13-17]). Although much of this research has focused on the technical aspects of development, rather than implementation and adoption [18], many of the themes that emerged from the World Café are echoed in the existing literature. Challenges introduced by the availability, quality, and standardization of data are consistently raised in discussions surrounding the use of AI in health care [19-21], and concerns about data privacy and security are also common [22,23]. Many studies investigating clinician uptake of AI-based CDSSs emphasize the importance of providing evidence-based decision support information [21,24-27], using development data that are reflective of the patient populations the AI is intended to support [28-30], and increasing the transparency of model reasoning as important facilitators to adoption [22-24,31]. These studies also underline the issues resulting from an absence of AI education in the current medical curricula and call for increased knowledge translation efforts to build trust and credibility with clinicians [24,32]. It is worth noting that several issues discussed by the participants, including time constraints and the need for intuitive and user-friendly applications, are pervasive issues in health care that are not necessarily unique to the adoption of AI-based tools [33,34]. Indeed, with a greater than 50% failure rate among many types of CDSS [35], the need for end user involvement in development has been highlighted as crucial to successful implementation [35,36].

While many of the issues identified in this study appear to be universal across health care, contextual idiosyncrasies remain. For example, the WCPs often stressed the importance of patient preference in CAD treatment decision-making and suggested that an AI-based CDSS may help facilitate patient discussions. In contrast, some prior studies have reported concerns about AI-based CDSSs negatively impacting the patient-clinician relationship, suggesting that using such systems could reduce the amount of time available for patient interaction or, in extreme cases, cut out contact altogether should clinical interactions be shifted to a digital format [22,25]. However, one of these studies interviewed only general practitioners [25], whose responsibilities are arguably more conducive to digitization than coronary revascularization decision-making. Similarly, other studies have discussed technical barriers beyond data requirements, including insufficient computing resources and inconsistent access to Wi-Fi, which are more pervasive in low-to-middle-income countries [28,37]. Thus, the importance of this work also lies in understanding the context within which the AI-based CDSS will be implemented, which is essential for successful adoption [38-41].

Finally, although AI-based clinical tools are increasingly being evaluated for their potential to enhance clinical decision-making and efficiency [42], their utility should be met with both optimism and caution. Recent evidence suggests that while AI-based systems can accurately perform specific clinical tasks and support information delivery, these capabilities may not necessarily translate into improved patient or system-level

XSL•FO
RenderX

outcomes [43,44]. Indeed, despite substantial investment and academic efforts, there remains limited prospective evidence demonstrating that AI-based tools have improved patient outcomes at scale [45]. Concerningly, among approved medical devices that use AI, clinical validation studies are inconsistently reported, and when reported, rarely include prospective and randomized evaluations [46,47]. Accordingly, while the technical performance of AI-based tools remains important, prospective randomized controlled trials are needed to demonstrate true clinical benefit. In parallel, future investigations should explicitly evaluate the fairness and equity of these tools and aim to determine whether such tools can alleviate clinician burden rather than add to cognitive or administrative loads [48].

We hope the diverse contextual factors described here can serve as a helpful foundation for the future development and implementation of CDSSs for CAD treatment decision-making.

Limitations

This study has several limitations. First, 3 clinician participants were unable to attend the World Café, and they were engaged in one-on-one interviews instead. Although efforts were made to engage these participants using comparable prompts and facilitation techniques, inherent differences between one-on-one and group-based World Café discussions may have influenced the data generated. As a result, some themes may reflect methodological differences despite efforts to reconcile findings across formats. Second, within our relatively small sample size, clinicians outnumbered the other stakeholder groups (8 vs 4), and their opinions may have dominated the discussions and results. For instance, some patient perspectives were articulated by clinicians rather than patients directly and may not necessarily represent the views or priorities of the patients themselves. Third, all participants were recruited in Alberta, limiting the generalizability of our findings to other jurisdictions. As the overall structure of AHS, including the clinical flow of patients and delivery of care, may not mirror that of other health care systems, some of the themes identified in the current investigation may be unique to this publicly funded and integrated health care system. Accordingly, while the themes and perspectives derived from this investigation provide valuable insight into the intricacies of implementing AI-based CDSS into practice, future work incorporating larger and more heterogeneous participant populations, balanced stakeholder representation, consistent data collection formats, and diverse health care system contexts may provide additional insights.

Conclusions

Various stakeholders, including patients and clinicians, believe that current coronary revascularization decision-making for patients with CAD is only partially evidence-based. AI-based CDSSs have the potential to improve this, leading to improved patient outcomes and health care cost savings. The successful development and implementation of such AI-based CDSSs hinges upon extensive end user involvement, data integration, data privacy protection, incorporation of patient preference, alignment with scientific evidence, and great usability. Integrating end user co-design and iterative usability testing may help support these priorities in future work.

Acknowledgments

We would like to thank the World Café participants for providing invaluable insights.

Funding

This study was supported by an AICE-Concepts Grant from Alberta Innovates (Accelerating Innovations Into Care; 212200473) and a project grant from the Canadian Institutes of Health Research (PJT 178027). The funders played no role in this study.

Authors' Contributions

KS designed and oversaw this study and wrote the manuscript. BB conducted this study, analyzed the data from the World Café and interviews, synthesized findings, and wrote the manuscript. CvR edited and revised the manuscript. BH and RW participated in the World Café and validated the main findings from this study. JL oversaw this study, provided resources, and prepared the manuscript. All authors proofread and approved the manuscript.

Conflicts of Interest

JL is a cofounder and major shareholder of Symbiotic AI, Inc. BH is a minor shareholder of Symbiotic AI, Inc. CvR is a paid independent contractor or consultant for Symbiotic AI, Inc. RW is an interventional cardiologist.

Checklist 1
COREQ checklist.
[PDF File, 432 KB - cardio_v10i1e81303_app1.pdf]

References

1. Gaudino M, Andreotti F, Kimura T. Current concepts in coronary artery revascularisation. Lancet 2023 May 13;401(10388):1611-1628. [doi: 10.1016/S0140-6736(23)00459-2] [Medline: 37121245]

XSL•FO
RenderX

2. Young MN, Kolte D, Cadigan ME, et al. Multidisciplinary Heart Team approach for complex coronary artery disease: single center clinical presentation. J Am Heart Assoc 2020 Apr 21;9(8):e014738. [doi: 10.1161/JAHA.119.014738] [Medline: 32308096]
3. Head SJ, Kaul S, Mack MJ, et al. The rationale for Heart Team decision-making for patients with stable, complex coronary artery disease. Eur Heart J 2013 Aug 21;34(32):2510-2518. [doi: 10.1093/eurheartj/eht059]
4. Walsh S, de Jong EEC, van Timmeren JE, et al. Decision support systems in oncology. JCO Clin Cancer Inform 2019 Feb;3:1-9. [doi: 10.1200/CCI.18.00001] [Medline: 30730766]
5. Löhr K, Weinhardt M, Sieber S. The "World Café" as a participatory method for collecting qualitative data. Int J Qual Methods 2020 Jan 1;19:1609406920916976. [doi: 10.1177/1609406920916976]
6. About AHS. Alberta Health Services. 2025 Dec 30. URL: https://www.albertahealthservices.ca/about/about.aspx [accessed 2026-01-21]
7. World Cafe method. World Cafe. 2020 Sep 11. URL: http://www.theworldcafe.com/key-concepts-resources/world-cafe-method [accessed 2026-01-21]
8. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
9. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches: Sage Publications; 2016.
10. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (Consolidated Criteria for Reporting Qualitative Studies). In: Guidelines for Reporting Health Research: A User's Manual: John Wiley & Sons, Ltd; 2014:214-226. [doi: 10.1002/9781118715598]
11. Bozyel S, Şimşek E, Koçyiğit Burunkaya D, et al. Artificial intelligence-based clinical decision support systems in cardiovascular diseases. Anatol J Cardiol 2024 Jan 7;28(2):74-86. [doi: 10.14744/AnatolJCardiol.2023.3685] [Medline: 38168009]
12. Gautam N, Saluja P, Malkawi A, et al. Current and future applications of artificial intelligence in coronary artery disease. Healthcare (Basel) 2022;10(2):232. [doi: 10.3390/healthcare10020232]
13. Bertsimas D, Orfanoudaki A, Weiner RB. Personalized treatment for coronary artery disease patients: a machine learning approach. Health Care Manag Sci 2020 Dec;23(4):482-506. [doi: 10.1007/s10729-020-09522-4] [Medline: 33040231]
14. Sajid M, Hassan A, Khan DA, et al. AI-CADS: an artificial intelligence based framework for automatic early detection and severity evaluation of coronary artery disease. Biomed Signal Process Control 2025 Aug;106:107705. [doi: 10.1016/j.bspc.2025.107705]
15. Yan J, Tian J, Yang H, et al. A clinical decision support system for predicting coronary artery stenosis in patients with suspected coronary heart disease. Comput Biol Med 2022 Dec;151(Pt A):106300. [doi: 10.1016/j.compbiomed.2022.106300] [Medline: 36410096]
16. Ninomiya K, Kageyama S, Shiomi H, et al. Can machine learning aid the selection of percutaneous vs surgical revascularization? J Am Coll Cardiol 2023 Nov 28;82(22):2113-2124. [doi: 10.1016/j.jacc.2023.09.818] [Medline: 37993203]
17. Kwon O, Na W, Kang H, et al. Electronic medical record-based machine learning approach to predict the risk of 30-day adverse cardiac events after invasive coronary treatment: machine learning model development and validation. JMIR Med Inform 2022 May 11;10(5):e26801. [doi: 10.2196/26801] [Medline: 35544292]
18. Bajgain B, Lorenzetti D, Lee J, Sauro K. Determinants of implementing artificial intelligence-based clinical decision support tools in healthcare: a scoping review protocol. BMJ Open 2023 Feb 23;13(2):e068373. [doi: 10.1136/bmjopen-2022-068373] [Medline: 36822813]
19. Bates DW, Auerbach A, Schulam P, Wright A, Saria S. Reporting and implementing interventions involving machine learning and artificial intelligence. Ann Intern Med 2020 Jun 2;172(11 Suppl):S137-S144. [doi: 10.7326/M19-0872] [Medline: 32479180]
20. Ji M, Chen X, Genchev GZ, Wei M, Yu G. Status of AI-enabled clinical decision support systems implementations in China. Methods Inf Med 2021 Dec;60(5-06):123-132. [doi: 10.1055/s-0041-1736461] [Medline: 34695871]
21. Kyrimi E, Dube K, Fenton N, et al. Bayesian networks in healthcare: what is preventing their adoption? Artif Intell Med 2021 Jun;116:102079. [doi: 10.1016/j.artmed.2021.102079] [Medline: 34020755]
22. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics 2021 Dec;47(12):e3. [doi: 10.1136/medethics-2019-105860]
23. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. Int J Med Inform 2022 May;161:104738. [doi: 10.1016/j.ijmedinf.2022.104738] [Medline: 35299098]
24. Abujaber AA, Nashwan AJ, Fadlalla A. Enabling the adoption of machine learning in clinical decision support: a total interpretive structural modeling approach. Inf Med Unlocked 2022;33:101090. [doi: 10.1016/j.imu.2022.101090]
25. Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T. General practitioners' attitudes toward artificial intelligence-enabled systems: interview study. J Med Internet Res 2022 Jan 27;24(1):e28916. [doi: 10.2196/28916] [Medline: 35084342]
26. Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. JAMIA Open 2022 Jul;5(2):ooac022. [doi: 10.1093/jamiaopen/ooac022] [Medline: 35474719]

XSL•FO
RenderX

27. van Gils AM, Visser LN, Hendriksen HM, et al. Assessing the views of professionals, patients, and care partners concerning the use of computer tools in memory clinics: international survey study. JMIR Form Res 2021 Dec 3;5(12):e31053. [doi: 10.2196/31053] [Medline: 34870612]
28. Emani S, Rui A, Rocha HAL, et al. Physicians' perceptions of and satisfaction with artificial intelligence in cancer treatment: a clinical decision support system experience and implications for low-middle-income countries. JMIR Cancer 2022 Apr 7;8(2):e31461. [doi: 10.2196/31461] [Medline: 35389353]
29. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. NPJ Digit Med 2022 Oct 28;5(1):162. [doi: 10.1038/s41746-022-00700-y] [Medline: 36307479]
30. Clement J, Maldonado AQ. Augmenting the transplant team with artificial intelligence: toward meaningful AI use in solid organ transplant. Front Immunol 2021;12:694222. [doi: 10.3389/fimmu.2021.694222] [Medline: 34177958]
31. Benza R, Kanwar M, Antaki J, et al. Effective communication of machine learning models in clinical decision support tools for PAH risk prediction. CHEST 2021 Oct;160(4):A1396-A1397. [doi: 10.1016/j.chest.2021.07.1277]
32. Goldstein MK, Coleman RW, Tu SW, et al. Translating research into practice: organizational issues in implementing automated decision support for hypertension in three medical centers. J Am Med Inform Assoc 2004;11(5):368-376. [doi: 10.1197/jamia.M1534] [Medline: 15187064]
33. Stagg BC, Stein JD, Medeiros FA, et al. Special commentary: using clinical decision support systems to bring predictive models to the glaucoma clinic. Ophthalmol Glaucoma 2021;4(1):5-9. [doi: 10.1016/j.ogla.2020.08.006] [Medline: 32810611]
34. Bentley KH, Zuromski KL, Fortgang RG, et al. Implementing machine learning models for suicide risk prediction in clinical practice: focus group study with hospital providers. JMIR Form Res 2022 Mar 11;6(3):e30946. [doi: 10.2196/30946] [Medline: 35275075]
35. Trivedi MH, Kern JK, Marcee A, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. Methods Inf Med 2002;41(5):435-442. [doi: 10.1055/s-0038-1634216] [Medline: 12501817]
36. Lindsell CJ, Stead WW, Johnson KB. Action-informed artificial intelligence-matching the algorithm to the problem. JAMA 2020 Jun 2;323(21):2141-2142. [doi: 10.1001/jama.2020.5035] [Medline: 32356878]
37. Jeong H, Kamaleswaran R. Pivotal challenges in artificial intelligence and machine learning applications for neonatal care. Semin Fetal Neonatal Med 2022 Oct;27(5):101393. [doi: 10.1016/j.siny.2022.101393] [Medline: 36266181]
38. Graham ID, Logan J, Harrison MB, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
39. McCormack B, Kitson A, Harvey G, Rycroft-Malone J, Titchen A, Seers K. Getting evidence into practice: the meaning of "context". J Adv Nurs 2002 Apr;38(1):94-104. [doi: 10.1046/j.1365-2648.2002.02150.x] [Medline: 11895535]
40. Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature. Milbank Q 2010 Dec;88(4):500-559. [doi: 10.1111/j.1468-0009.2010.00611.x] [Medline: 21166868]
41. Tomoaia-Cotisel A, Scammon DL, Waitzman NJ, et al. Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change. Ann Fam Med 2013;11 Suppl 1(Suppl 1):S115-S123. [doi: 10.1370/afm.1549] [Medline: 23690380]
42. Khan E, Lambrakis K, Briffa T, et al. Re-engineering the clinical approach to suspected cardiac chest pain assessment in the emergency department by expediting research evidence to practice using artificial intelligence. (RAPIDx AI)-a cluster randomized study design. Am Heart J 2025 Jul;285:106-118. [doi: 10.1016/j.ahj.2025.02.016] [Medline: 39993551]
43. Mazor T, Farhat KS, Trukhanov P, et al. Clinical trial notifications triggered by artificial intelligence-detected cancer progression: a randomized trial. JAMA Netw Open 2025 Apr 1;8(4):e252013. [doi: 10.1001/jamanetworkopen.2025.2013] [Medline: 40257799]
44. Mandair D, Elia MV, Hong JC. Considerations in translating AI to improve care. JAMA Netw Open 2025 Apr 1;8(4):e252023. [doi: 10.1001/jamanetworkopen.2025.2023] [Medline: 40257804]
45. Armoundas AA, Narayan SM, Arnett DK, et al. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. Circulation 2024 Apr 2;149(14):e1028-e1050. [doi: 10.1161/CIR.0000000000001201] [Medline: 38415358]
46. El Fassi SC, Abdullah A, Fang Y, et al. Not all AI health tools with regulatory authorization are clinically validated. Nat Med 2024 Oct;30(10):2718-2720. [doi: 10.1038/s41591-024-03203-3] [Medline: 39187696]
47. Windecker D, Baj G, Shiri I, et al. Generalizability of FDA-approved AI-enabled medical devices for clinical use. JAMA Netw Open 2025 Apr 1;8(4):e258052. [doi: 10.1001/jamanetworkopen.2025.8052] [Medline: 40305017]
48. Khera R, Butte AJ, Berkwits M, et al. AI in medicine-JAMA's focus on clinical outcomes, patient-centered care, quality, and equity. JAMA 2023 Sep 5;330(9):818-820. [doi: 10.1001/jama.2023.15481] [Medline: 37566406]

Abbreviations

AHS: Alberta Health Services
AI: artificial intelligence
CABG: coronary artery bypass grafting

XSL•FO
RenderX

CAD: coronary artery disease
CDSS: clinical decision support system
COREQ: Consolidated Criteria for Reporting Qualitative Research
OIP: one-on-one interview participant
WCP: World Café participant

Edited by KC Wong; submitted 25.Jul.2025; peer-reviewed by AA Arafat, D Yoo, K Johl; revised version received 31.Dec.2025; accepted 11.Jan.2026; published 06.Feb.2026.

Please cite as:
Sauro K, Bajgain B, van Rassel C, Har B, Welsh R, Lee J
Perceived Potential and Challenges of Supporting Coronary Artery Disease Treatment Decisions With AI: Qualitative Study
JMIR Cardio 2026;10:e81303
URL: https://cardio.jmir.org/2026/1/e81303
doi:10.2196/81303

© Khara Sauro, Bishnu Bajgain, Cody van Rassel, Bryan Har, Robert Welsh, Joon Lee. Originally published in JMIR Cardio (https://cardio.jmir.org), 6.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a Cloud-Based Care Coordination Platform on False Cardiac Catheterization Laboratory Activations and Unnecessary Team Mobilization: Retrospective Cohort Study

William Gibson[1*], MBChB; Dawoud Al Kindi[1*], MD; François Brouillette[1*], MD; Ahmed Aldajani[2], MD; Omar Chaabo[1], MD; Yasmine Lachance[3], MD; Elie Akl[1], MD; Kshitij Badal Dandona[1], MD; Giuseppe Martucci[1], MD; Jean-Philippe Pelletier[1], MD; Nicolo Piazza[1], MD, PhD; Jeremy Y Levett[1], MD; Tomer Moran[4], BSc; Marco Spaziano[1], MD, MSc

[1]Department of Cardiology, McGill University Health Centre, 1001 Decarie Blvd, Montréal, QC, Canada

[2]Department of Internal Medicine, Division of Cardiology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

[3]Department of Cardiology, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada

[4]Department of Cardiology, McGill University, Montréal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
William Gibson, MBChB
Department of Cardiology, McGill University Health Centre, 1001 Decarie Blvd, Montréal, QC, Canada

Abstract

Background: Rapid activation of the cardiac catheterization laboratory (CCL) for ST-segment elevation myocardial infarction (STEMI) is essential to minimize time to reperfusion. However, system-wide efforts to reduce treatment delays have been accompanied by increased false activations, defined as activations that do not result in emergent coronary intervention. False activations contribute to unnecessary team mobilization (UTM), staff fatigue, workflow disruption, and inefficient resource use.

Objective: This study aimed to evaluate whether the implementation of a cloud-based care coordination platform (Stenoa) was associated with reductions in false activations and UTMs at a high-volume tertiary cardiac center.

Methods: In September 2021, the McGill University Health Centre implemented Stenoa, a mobile, cloud-based STEMI coordination platform enabling systematic case validation using electrocardiographic and clinical data. A retrospective cohort study was conducted, including all CCL activations between September 2020 and December 2022. Activations were grouped as preimplementation (group 0: September 2020 to September 2021) and postimplementation (group 1: September 2021 to December 2022) periods. A false activation was defined as a CCL activation followed by case cancellation before any procedure was performed. The primary outcome was the rate of UTM.

Results: In total, 632 activations were analyzed (group 0: n=288; group 1: n=344). UTM decreased from 8.7% (23/265) to 4.4% (14/316) following platform implementation (P=.04). False activation frequency decreased from 10.2% (27/265) to 6.9% (22/316), although this difference did not reach statistical significance (P=.16). Among false activations, the proportion resulting in UTM declined from 85% to 63% (P=.08).

Conclusions: The implementation of a cloud-based STEMI coordination platform was associated with a significant reduction in unnecessary catheterization laboratory team mobilization. Structured digital communication may improve workflow efficiency and resource use in STEMI systems of care. Further multicenter evaluation is warranted.

(JMIR Cardio 2026;10:e76932) doi:10.2196/76932

KEYWORDS

cloud-based app; ST-elevated myocardial infarction; STEMI; false activation; clinical decision support; digital health; smartphone app; acute coronary syndrome; telemedicine; mobile health; catheterization laboratory; reperfusion time; digital communication; quality improvement; resource utilization

Introduction

Every minute of ischemia counts in acute coronary syndrome (ACS), and improving time to reperfusion has become a core priority of health systems worldwide [1]. Current guidelines recommend primary percutaneous coronary intervention within 90 to 120 minutes (depending on the initial presenting location) from first medical contact (FMC) to coronary blood flow restoration as a system goal [2]. Over the past 2 decades, concerted system-level interventions have led to substantial

XSL•FO
RenderX

reductions in reperfusion delay, an achievement accompanied by corresponding declines in in-hospital mortality [3,4]. These advances reflect the success of coordinated, protocol-driven pathways that emphasize early recognition, prehospital electrocardiogram (ECG) acquisition, and direct activation of the cardiac catheterization laboratory (CCL) [5]. Efficient ST-segment elevation myocardial infarction (STEMI) care depends on the seamless integration of 2 interdependent components: the STEMI procedure**,** referring to definitive revascularization via primary percutaneous coronary intervention once the patient reaches the CCL, and the STEMI process, which encompasses the upstream sequence of assessments, communications, and logistical actions that move patients from symptom onset through evaluation, triage, transport, and activation. At the center of this process lies accurate ECG-based identification of true STEMI, but fewer than 5% of chest pain presentations represent true STEMI, and while essential to STEMI diagnosis, ECG interpretation is imperfect, and several mimicking conditions can lead to unnecessary CCL activation if misidentified. As such, systems must balance rapid activation with diagnostic accuracy to deliver timely care while avoiding overtriage.

As systems have optimized speed and broadened access through earlier upstream activation pathways, a parallel concern has emerged: an increase in false or inappropriate activations, highlighting the need to achieve timeliness while maintaining diagnostic specificity. False activations—defined as CCL activations for patients who ultimately do not require emergent coronary angiography—have gained increasing recognition as an unintended by-product of system-wide efforts to shorten reperfusion times. Reported rates vary widely, from approximately 5% to 40%, reflecting heterogeneity in definitions, emergency medical service training, and activation workflows across regions [6]. Their impact is considerable: unnecessary team mobilizations (UTMs) contribute to staff fatigue and resource strain, impose avoidable financial costs [6], and may undermine patient confidence when activations are later deemed unwarranted. As health systems continue to prioritize timeliness, preserving diagnostic precision while minimizing inappropriate activations has become an important quality challenge in contemporary STEMI care.

Recently, mobile apps, telemedicine, and cloud-based platforms have markedly improved communication and coordination within the emergency care setting [7,8]. Digital transmission of prehospital ECGs to interventional cardiologists enables earlier diagnosis, triage, and direct activation of the CCL—streamlining workflow and reducing reperfusion delay. In a recent systematic review and meta-analysis of 17 studies including >4300 patients, prehospital digital ECG transmission was associated with significant reductions in door-to-device time (mean difference −33 minutes, 95% CI −50.5 to −16.2), FMC-to-device time (mean difference −24.7 minutes, 95% CI −37.1 to −12.3), and a 47% lower mortality compared with standard care (odds ratio 0.53, 95% CI 0.40 - 0.69) [9]. Similarly, a recent meta-analysis of smartphone-based STEMI coordination apps demonstrated consistent time-efficiency benefits, with pooled reductions of approximately 19 minutes in door-to-balloon time and 20 minutes in FMC-to-balloon time compared with conventional activation methods, without compromising the accuracy of triage or procedural outcomes [10]. Although false activation rates could not be pooled due to variability in definitions, several included studies suggested a trend toward fewer inappropriate activations with structured app-based communication. These findings highlight the growing role of digital health tools in optimizing STEMI pathways and suggest that standardized, platform-based coordination may enhance both timeliness and accuracy of CCL activation.

Stenoa is a cloud-based mobile app dedicated to improving the CCL activation pathway by creating a single direct communication channel between the various members of the chain of survival (paramedics, emergency physicians, interventional cardiologists, and CCL teams), rather than using traditional means of communication (telephone, locating, fax, email, and SMS). Previous work from our center showed that implementation of the Stenoa platform was associated with reduced door-to-balloon times [11]. The aim of this study was to evaluate the impact of Stenoa implementation on false activation rates and the resultant effect on UTM at a single center.

Methods

Study Design and Setting

This retrospective observational study was conducted and reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. The completed STROBE checklist is included as Multimedia Appendix 1. This study evaluated all consecutive CCL activations at the McGill University Health Centre (MUHC) from September 2020 to December 2022, comparing outcomes before and after implementation of the mobile app. MUHC is a tertiary referral center at which personnel are off-site outside of regular working hours. Included patients were aged ≥18 years with a suspected unstable cardiac condition requiring emergent activation of the CCL. STEMI was the primary indication for CCL activation. Other indications for which the CCL was solicited included high-risk non-STEMI (NSTEMI), temporary transvenous pacemaker insertion, mechanical circulatory support for cardiogenic shock, and pericardiocentesis for cardiac tamponade. Before Stenoa implementation, CCL activation (and deactivation) was undertaken by the interventional cardiologist or emergency physician through the hospital locating services using a paging system. ECGs were either described over the phone or shared via SMS. After Stenoa implementation, CCL activation was done exclusively by the interventional cardiologist through the platform after clinical and electrocardiographic data were submitted and assessed. Similarly, prompt case deactivation could be achieved through the platform.

Stenoa Workflow

The Stenoa platform is a secure, cloud-based mobile app designed to streamline communication and decision-making during suspected ACS presentations requiring potential CCL activation (Figure 1). Patient encounters originate either from direct emergency department presentation or from the community via paramedic services. The FMC provider performs

XSL•FO
RenderX

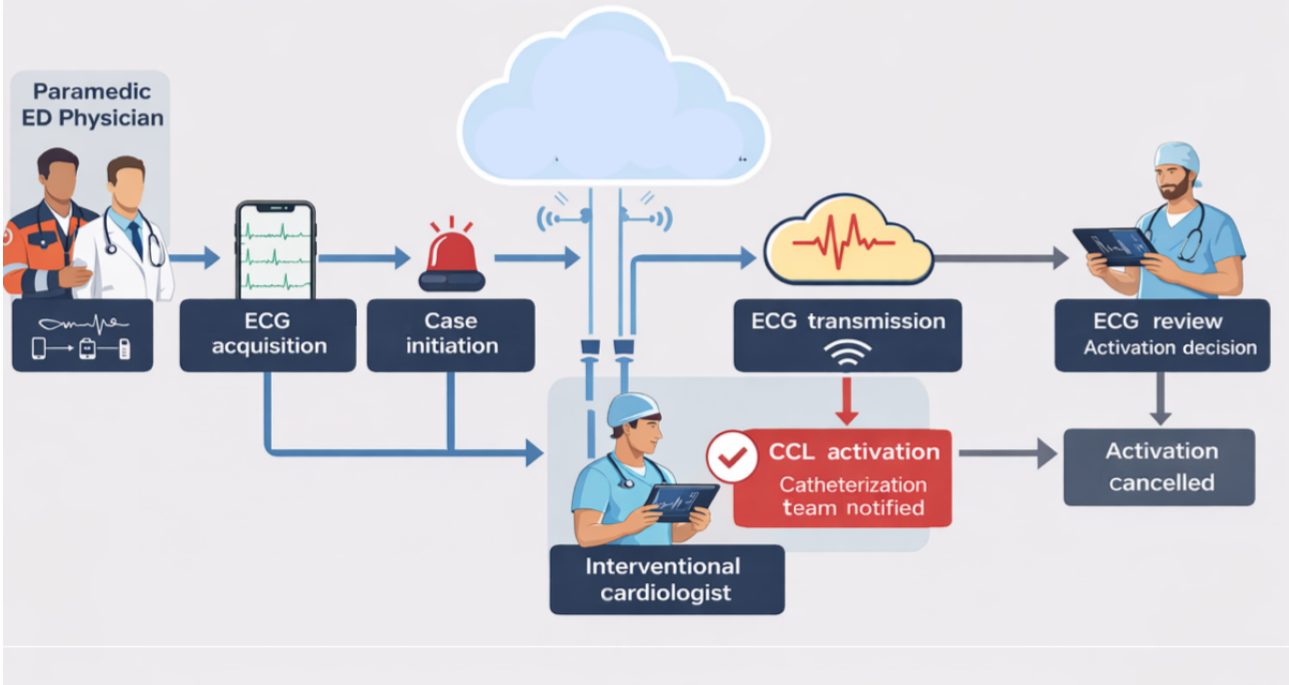
the initial clinical assessment based on symptoms, physical examination, and 12-lead ECG. When STEMI or another emergent cardiac condition warranting urgent CCL activation is suspected, a case is initiated within the Stenoa app, including patient demographics, clinical details, and ECG tracing. All uploaded information is fully encrypted and compliant with the Health Insurance Portability and Accountability Act (HIPAA) and the Systems and Organization Controls 2 security standards.

Upon case creation, coordinated real-time alerts are automatically issued to the on-call interventional cardiologist and CCL team. The interventional cardiologist reviews the submitted information through the platform and determines whether immediate activation of the CCL is appropriate. Secure messaging and integrated voice communication allow bidirectional discussion between the referring clinician and interventional cardiologist to clarify diagnostic uncertainty, if required. The final decision—activation or cancellation—is entered directly into the platform, prompting automated notification to all relevant team members (nursing staff, technicians, catheterization laboratory coordinator, and anesthesia team, if applicable).

In addition to facilitating the clinical workflow, the platform automatically captures key system performance time interval metrics, including FMC-to-ECG time, ECG-to-CCL activation time, FMC-to-balloon time, and CCL activation-to-balloon time. These time stamps are stored in real time to support internal auditing, quality assessment, and performance benchmarking.

Figure 1 presents a schematic representation of the Stenoa cloud-based communication platform used for suspected ACS requiring potential CCL activation. Following FMC, clinicians initiate a case within the platform and upload electrocardiographic data and clinical information. The on-call interventional cardiologist reviews the case in real time and determines whether CCL activation or cancellation is appropriate, with automated notifications sent to the catheterization team.

Figure 1. Stenoa platform workflow for cardiac catheterization laboratory activation. CCL: cardiac catheterization laboratory; ECG: electrocardiogram; ED: emergency department.



Definitions

False activation was defined as activation of the CCL team for which no procedure was performed. UTM was defined as any false activation case that resulted in team mobilization to the hospital. False activation rates and the proportion resulting in UTM were compared before and after implementation of the app. Final diagnoses were determined based on the treating physician's documentation in the medical record following clinical evaluation and catheterization findings. The primary outcome of this study was the overall rate of UTM.

Participants

Patients were divided into 2 groups for analysis. Group 0 consisted of all CCL activations during the 12 months before the implementation of Stenoa (September 1, 2020, to August 31, 2021). Group 1 consisted of all CCL activations during the 15 months immediately after Stenoa's introduction. The first 3 months (September 1, 2021, to November 30, 2021) were a transition period in which both systems were used. During the following 12 months (December 1, 2021, to December 31, 2022), CCL activations (and deactivations) were done exclusively via the platform.

Data were obtained from institutional medical records and the Stenoa platform database. Chart review was authorized by the Direction of Professional Services of the MUHC as part of a quality and outcomes evaluation.

UTM events were identified using institutional overtime billing records. These records provide an objective and auditable measure of CCL team mobilization, as clinical personnel are compensated on a per-case basis when physically mobilized to

XSL•FO
RenderX

the hospital outside scheduled hours. Because CCL staff are typically mobilized from home during off-hours, overtime billing logs serve as a reliable surrogate for off-hours team mobilization events. Accordingly, the UTM metric reflects the operational burden of false activations requiring off-hours staff mobilization rather than all canceled activations occurring across the full 24-hour workflow. False activations that occurred while the CCL team was already on site (ie, during weekdays or because of a previous case) or false activations that were deactivated within 15 minutes of activation were not considered to have led to UTM. Accordingly, overtime billing logs were used as the most objective and auditable measure of team mobilization in cases of false or canceled activations.

Ethical Considerations

This study was conducted as a retrospective observational analysis using anonymized clinical and operational data obtained from institutional medical records and the Stenoa platform. No direct patient contact or intervention occurred as part of the study. In accordance with institutional policies [12] governing retrospective analyses using deidentified data, formal research ethics board review and individual informed consent requirements were waived by the MUHC. All data were deidentified prior to analysis and handled in accordance with institutional privacy and confidentiality policies. No identifiable patient images are included in this manuscript or supplementary materials. Participants received no compensation, as the study involved retrospective analysis of existing clinical records.

Statistical Methods

Categorical variables were summarized as frequencies and percentages, and between-group differences were evaluated using chi-square tests. Continuous variables were assessed for normality and presented as mean (SD) or median (IQR), as appropriate. A 2-sided *P* value <.05 was considered statistically significant. All analyses were conducted using standard statistical software. Data were recorded prospectively in a secure institutional database and subsequently exported for analysis.

Results

In total, 632 patients were included in this study: 288 (45.6%) in group 0 (preimplementation group) and 344 (54.4%) in group 1 (postimplementation group). Of these, 92% (265/288 in group 0 and 316/344 in group 1) of activations occurred off-hours in both groups. The overall rate of UTM declined from 8.7% (23/265) in group 0 to 4.4% (14/316) in group 1 (*P*=.04). In total, 27 of 265 (10.2%) false activations occurred in group 0, compared with 22 of 316 (7%) false activations in group 1 (*P*=.16; Table 1). In group 0, 23 of 27 (85.2%) false activations resulted in UTM, compared with 14 of 22 (63.6%) false activations in group 1 (*P*=.08). Detailed demographic data were not available for the full study cohort; age and sex data were available only among patients with false activations. Among patients with false activations, those in group 1 were older on average (mean 71.1, SD 16.1 years) than those in group 0 (mean 67.5, SD 17.0 years), although this difference was not statistically significant (*P*=.44). In addition, the proportion of male patients was higher in group 1 (18/22, 81.8%) than that in group 0 (15/27, 55.6%), whereas female representation was greater in group 1 (15/27, 56%; *P*=.02).

Table . Rates of false activations and unnecessary team mobilization.

Outcome	Group 0[a]	Group 1[b]	*P* value
Unnecessary team mobilization, n (%)	23 (8.7)	14 (4.4)	.04
False activations, n (%)	27 (10.2)	22 (7)	.16
Unnecessary team mobilization among false activations (%)[c]	85.2	63.6	.08

[a]Group 0: n=265.

[b]Group 1: n=316.

[c]Percentages are calculated as the proportion of false activations that resulted in unnecessary team mobilization.

Reasons for false activations are shown in Table 2 and were similar between groups (*P*>.05). Across both cohorts, the most common cause was an alternative diagnosis other than ACS (26/49, 53.1%), including pericarditis, arrhythmias, and old myocardial infarction with persistent ST elevation. Other ACS presentations not requiring immediate angiography (NSTEMI or unstable angina) accounted for 30.6% (15/49), and factors related to patient suitability (eg, dementia, advanced comorbidity, and palliative care) accounted for 12.2% (6/49). Additionally, 1 of 49 (2%) cases was redirected to another CCL, and another involved a patient presenting with recurrent chest pain and equivocal ECG changes 4 days after an index infarction. This patient was subsequently found to have ongoing ischemia requiring repeat intervention. This event reflected a delayed presentation outside the typical primary PCI window rather than a missed activation attributable to the coordination platform. Reasons for UTM were similarly analyzed. In group 0, 39.1% (9/23) of UTM cases were unrelated to ACS, 43.5% (10/23) involved ACS presentations that did not require emergent angiography, and 17.9% (4/23) involved patients with advanced cancer receiving palliative care. In group 1, 57.1% (8/14) of UTM cases were not ACS related, 28.6% (4/14) were ACS presentations requiring delayed angiography, and 14.3% (2/14) involved prohibitive comorbidities or palliative status.

XSL•FO
RenderX

Table . Causes of false activations before and after implementation.

Cause of false activation	Group 0 (before Stenoa)[a]	Group 1 (after Stenoa)[b]
Alternative diagnosis (non-ACS[c]), n (%)	13 (48.1)	13 (59.1)
Other ACS not requiring emergent angiography, n (%)	10 (37)	5 (22.7)
Palliative care or advanced malignancy, n (%)	4 (14.8)	0 (0)
Comorbidities, n (%)	0 (0)	2 (9.1)
Late-presenting myocardial infarction or redirected case, n (%)	0 (0)	2 (9.1)

[a]Group 0: n=27.

[b]Group 1: n=22.

[c]ACS: acute coronary syndrome.

Discussion

Principal Findings

This study demonstrates a significant reduction in UTM following implementation of a cloud-based care coordination platform, decreasing from 23 (8.7%) cases before implementation to 14 (4.4%) cases after implementation (*P*=.04). Although the reduction in false activations did not reach statistical significance, the proportion of false activations that resulted in UTM decreased meaningfully from 85% to 64%, indicating more precise activation decision-making.

These improvements may be explained by the mechanisms inherent to the platform workflow. First, Stenoa increases the amount of diagnostic and contextual information available to the interventional cardiologist prior to activation, reducing reliance on incomplete early data and minimizing "blind" activations. Second, real-time secure communication enables rapid case reassessment and deactivation in response to evolving clinical information, preventing unnecessary mobilization even when activation is initially triggered. Third, structured team communication eliminates fragmented parallel notification pathways that often delay clarification or cancellation in conventional systems. Consequently, cases previously likely to progress to full mobilization despite diagnostic uncertainty may now be clarified earlier in the decision chain, lowering UTM burden without sacrificing sensitivity. Importantly, the reduction in UTM should be interpreted as an operational improvement rather than solely as a change in diagnostic accuracy. In conventional paging systems, once activation is initiated, cancellation frequently occurs after staff have already begun mobilization. By enabling earlier clarification of borderline cases, structured digital communication may prevent unnecessary mobilization before staff departure, thereby reducing operational burden even when the initial activation threshold remains unchanged. Furthermore, no cases of delayed STEMI recognition attributable to the activation workflow were identified during the study period.

The reported incidence of false activations varies widely across STEMI systems, largely due to differing definitions, diagnostic criteria, and activation workflows used in prior studies. Early prospective registry data from Larson et al [13] demonstrated a false activation rate of 14% among 1345 consecutive activations within a regional STEMI network; however, after excluding patients with ST elevation and positive biomarkers (who meet the universal STEMI definition despite the absence of an angiographic culprit), the true false activation rate was closer to 9.2%. In contrast, Youngquist et al [14] found substantially higher false activation rates with prehospital ECG–based activation than with emergency-physician activation (39% vs 9%; *P*=.02), although this was limited by small sample size and reliance on automated ECG interpretation alone. Other centers report rates ranging from 11% to 15% using coronary angiography–based definitions, such as the emergency-physician activation cohort in the study by Kontos et al [15] (14.9% false positive activations) [16] and the registry of 489 activations by Nfor et al [17] (11% without culprit lesion). McCabe et al [18] reported a higher false activation rate of 36% in the activate-SF registry using composite criteria integrating angiographic, clinical, and biomarker data; however, their definition classified patients with elevated biomarkers but without an angiographic culprit as false positives—despite technically fulfilling universal STEMI diagnostic criteria [16]—illustrating the degree of variability created by variations in the definitions adopted [18]. Similarly, statewide system data from the RACE program demonstrated inappropriate activation rates between 15% and 18.7%, depending on classification methodology [19]. Collectively, these studies highlight the absence of standardized false activation definitions and demonstrate that reported rates may differ 3-fold depending on the operational framework used, underscoring the challenge of benchmarking performance across systems. Meanwhile, the false activation rate in our center (9.4% preimplementation group) lies at the lower end of the published range. This likely reflects both definitional differences and the inclusion of high-risk NSTEMI presentations in our center, an approach that aligns with contemporary thinking around acute coronary occlusion. Growing evidence demonstrates that a substantial proportion of patients with acute coronary occlusion do not meet classic STEMI ECG criteria yet experience comparable infarct burden and benefit from urgent reperfusion. The emerging occlusive myocardial infarction model suggests that relying solely on ST-segment elevation thresholds likely risks delaying treatment among a significant cohort of patients with true coronary occlusion. In this context, many cases traditionally labeled as "inappropriate" or "false" activations may instead represent clinically justified attempts to avoid

XSL•FO
RenderX

missed occlusive myocardial infarction [20]. These realities emphasize the importance of activation processes that allow clinicians to apply judgment and communicate rapidly when uncertainty exists, rather than depending solely on binary ECG criteria. Workflow solutions that enhance such collaboration—like our platform—may therefore contribute to more appropriate CCL activation.

Limitations

Clinical Limitations

First, this was a single-center study, which may limit generalizability. Second, the baseline false activation rate in our institution was relatively low (approximately 10%), reducing statistical power to detect differences in false activation incidence. We applied a clinically oriented definition of false activation that did not classify cases with ECG findings warranting emergent angiography but without an occluded vessel (eg, Takotsubo cardiomyopathy) as false activations. Although some prior publications categorize such cases as false activations, doing so retrospectively risks misclassification bias, as these patients typically present with symptoms and ECG features indistinguishable from true STEMI at the point of decision-making. Third, our study does not include all chest pain evaluations in the emergency care pathway; therefore, we could not assess false-negative activations (ie, missed STEMI) in the broader population. Fourth, catheterization laboratory staff experience and satisfaction were not assessed, and future work may explore operational impact beyond mobilization frequency. Fifth, this study lacks complete demographic data for the overall cohort, as age and sex information were only available for patients with false activations. Consequently, baseline comparability between the full preimplementation and postimplementation groups could not be formally assessed. Within the false activation subgroup, differences in sex distribution and a nonsignificant difference in age were observed, suggesting potential population heterogeneity. However, given the small size of this subgroup and the system-level nature of the intervention, these differences are unlikely to fully account for the observed reduction in UTM. Nevertheless, residual confounding cannot be excluded. Finally, the study period overlapped with the COVID-19 pandemic; however, review of each false activation case did not identify instances in which pandemic-associated ECG abnormalities influenced activation decisions.

Operational Limitations

Implementation of the platform coincided with a change in activation workflow whereby final activation authority was centralized to the interventional cardiologist rather than shared between emergency physicians and cardiologists as in the previous system. As such, it is not possible to fully separate the effects of the digital platform itself from the impact of this workflow modification. Furthermore, detailed workflow timing metrics, such as ECG-to-activation decision time, were not available within the present dataset. However, prior work from our center evaluating the Stenoa platform demonstrated significant reductions in door-to-balloon time without delays in activation decision-making [11]. The definition of UTM was based on overtime billing records, which accurately capture episodes of team mobilization but do not account for situations in which staff were mobilized multiple times within short intervals or were delayed from leaving the hospital, potentially underestimating the practical burden on personnel and workflow. In addition, causes of false activation were determined from treating physician documentation in unblinded clinical records rather than by independent blinded adjudication, which may introduce observer bias and potential misclassification. This method of assessing UTM primarily captures off-hours mobilization burden and may not fully represent cancellations occurring during regular staffed hours. Additionally, substantial heterogeneity exists across centers in false activation definitions and activation criteria, influenced by differences in local protocols, patient populations, and operational thresholds. These challenges underscore the current need for standardized terminology and reporting frameworks to enable more meaningful comparison across systems [21]. In addition, the first 3 months following platform implementation represented a transition period during which both the traditional paging system and the Stenoa platform were used. Because individual cases could not be reliably attributed to a single system during this interval, sensitivity analyses excluding the transition period could not be performed. This may introduce a small degree of misclassification when comparing preimplementation and postimplementation periods. In this study, the consideration of UTM was grounded in resource use: if the CCL team was mobilized for a case that ultimately did not result in emergent intervention, the activation was considered unnecessary from a system perspective. Although cost-benefit analysis was beyond the scope of this manuscript, a reduction in UTMs could result in lower resource use over time.

Conclusions

Implementation of a cloud-based STEMI coordination platform was associated with a significant reduction in unnecessary catheterization laboratory team mobilization. Therefore, digital coordination tools may represent an effective strategy for improving workflow efficiency and resource use in STEMI systems of care.

Acknowledgments

The authors acknowledge the cardiac catheterization laboratory staff, emergency department personnel, and paramedic teams at the McGill University Health Centre for their contributions to patient care and data collection.

Generative artificial intelligence tools were used for language editing and refinement of the manuscript. All scientific content, study design, data analysis, and interpretations were developed by the authors. The authors take full responsibility for the integrity and accuracy of the study.

XSL•FO
RenderX

Funding

The authors declare that no financial support was received for this study.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to institutional data protection policies but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MS, DAK, WG
Data curation: MS
Formal analysis: MS, DAK
Methodology: MS, DAK
Supervision: MS
Writing—original draft: DAK
Writing—review and editing: MS, WG, FB, AA, OC, YL, EA, GM, JPP, NP, KBD, JYL, TM

Conflicts of Interest

MS serves as a medical officer for Stenoa and holds equity in the company. JYL holds a leadership position at Stenoa and holds equity in the company. TM has equity in Stenoa. All other authors have no other conflicts of interest to declare.

Multimedia Appendix 1
STROBE checklist.
[PDF File, 84 KB - cardio_v10i1e76932_app1.pdf]

References

1. Wong GC, Welsford M, Ainsworth C, et al. 2019 Canadian Cardiovascular Society/Canadian Association of Interventional Cardiology Guidelines on the acute management of ST-elevation myocardial infarction: focused update on regionalization and reperfusion. Can J Cardiol 2019 Feb;35(2):107-132. [doi: 10.1016/j.cjca.2018.11.031] [Medline: 30760415]
2. Byrne RA, Rossello X, Coughlan JJ, et al. 2023 ESC guidelines for the management of acute coronary syndromes [Article in Italian]. G Ital Cardiol (Rome) 2024 Feb;25(2 Suppl 2):e1-e112. [doi: 10.1714/4191.41785] [Medline: 38291910]
3. Krumholz HM, Herrin J, Miller LE, et al. Improvements in door-to-balloon time in the United States, 2005 to 2010. Circulation 2011 Aug 30;124(9):1038-1045. [doi: 10.1161/CIRCULATIONAHA.111.044107] [Medline: 21859971]
4. Foo CY, Bonsu KO, Nallamothu BK, et al. Coronary intervention door-to-balloon time and outcomes in ST-elevation myocardial infarction: a meta-analysis. Heart 2018 Aug;104(16):1362-1369. [doi: 10.1136/heartjnl-2017-312517] [Medline: 29437704]
5. Brunetti ND, De Gennaro L, Correale M, et al. Pre-hospital electrocardiogram triage with telemedicine near halves time to treatment in STEMI: a meta-analysis and meta-regression analysis of non-randomized studies. Int J Cardiol 2017 Apr 1;232:5-11. [doi: 10.1016/j.ijcard.2017.01.055] [Medline: 28089154]
6. Henry TD, Younger L, Derakhshan A, et al. Economic impact of false ST-segment elevation myocardial infarction (STEMI) cardiac catheterization laboratory (CCL) activations at a major Los Angeles County STEMI-receiving center (SRC). J Am Coll Cardiol 2016;67(13_Supplement). [doi: 10.1016/S0735-1097(16)30636-2]
7. Plaza Roncero A, Marques G, Sainz-De-Abajo B, et al. Mobile health apps for medical emergencies: systematic review. JMIR Mhealth Uhealth 2020 Dec 11;8(12):e18513. [doi: 10.2196/18513] [Medline: 33306037]
8. Marcolino MS, Maia LM, Oliveira JA, et al. Impact of telemedicine interventions on mortality in patients with acute myocardial infarction: a systematic review and meta-analysis. Heart 2019 Oct;105(19):1479-1486. [doi: 10.1136/heartjnl-2018-314539] [Medline: 31253696]
9. Moxham RN, d'Entremont MA, Mir H, Schwalm JD, Natarajan MK, Jolly SS. Effect of prehospital digital electrocardiogram transmission on revascularization delays and mortality in ST-elevation myocardial infarction patients: systematic review and meta-analysis. CJC Open 2024;6(10):1199-1206. [doi: 10.1016/j.cjco.2024.06.012] [Medline: 39525334]
10. Gibson W, Al Kindi D, Akl E, et al. Impact of smartphone apps on reperfusion times and clinical outcomes in acute ST-segment elevation myocardial infarction: systematic review and meta-analysis. JMIR Mhealth Uhealth 2025 Aug 25;13:e66605. [doi: 10.2196/66605] [Medline: 40854299]
11. Aldajani A, Chaabo O, Brouillette F, et al. Impact of a mobile, cloud-based care-coordination platform on door-to-balloon time in patients with STEMI: initial results. CJC Open 2024;6(9):1094-1098. [doi: 10.1016/j.cjco.2024.05.011] [Medline: 39525824]
12. Tri-council policy statement: ethical conduct for research involving humans. Government of Canada. 2018. URL: https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html [accessed 2025-05-25]

XSL•FO
RenderX

13. Larson DM, Menssen KM, Sharkey SW, et al. “False-positive” cardiac catheterization laboratory activation among patients with suspected ST-segment elevation myocardial infarction. JAMA 2007 Dec 19;298(23):2754-2760. [doi: 10.1001/jama.298.23.2754] [Medline: 18165668]
14. Youngquist ST, Shah AP, Niemann JT, Kaji AH, French WJ. A comparison of door-to-balloon times and false-positive activations between emergency department and out-of-hospital activation of the coronary catheterization team. Acad Emerg Med 2008 Aug;15(8):784-787 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00186.x] [Medline: 18638030]
15. Kontos MC, Kurz MC, Roberts CS, et al. An evaluation of the accuracy of emergency physician activation of the cardiac catheterization laboratory for patients with suspected ST-segment elevation myocardial infarction. Ann Emerg Med 2010 May;55(5):423-430. [doi: 10.1016/j.annemergmed.2009.08.011] [Medline: 19767126]
16. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). Circulation 2018 Nov 13;138(20):e618-e651. [doi: 10.1161/CIR.0000000000000617] [Medline: 30571511]
17. Nfor T, Kostopoulos L, Hashim H, et al. Identifying false-positive ST-elevation myocardial infarction in emergency department patients. J Emerg Med 2012 Oct;43(4):561-567. [doi: 10.1016/j.jemermed.2011.09.027] [Medline: 22284976]
18. McCabe JM, Armstrong EJ, Kulkarni A, et al. Prevalence and factors associated with false-positive ST-segment elevation myocardial infarction diagnoses at primary percutaneous coronary intervention–capable centers: a report from the Activate-SF registry. Arch Intern Med 2012 Jun 11;172(11):864-871. [doi: 10.1001/archinternmed.2012.945] [Medline: 22566489]
19. Garvey JL, Monk L, Granger CB, et al. Rates of cardiac catheterization cancelation for ST-segment elevation myocardial infarction after activation by emergency medical services or emergency physicians: results from the North Carolina Catheterization Laboratory Activation Registry. Circulation 2012 Jan 17;125(2):308-313. [doi: 10.1161/CIRCULATIONAHA.110.007039] [Medline: 22147904]
20. McLaren J, de Alencar JN, Aslanger EK, Meyers HP, Smith SW. From ST-segment elevation MI to occlusion MI: the new paradigm shift in acute myocardial infarction. JACC Adv 2024;3(11):101314. [doi: 10.1016/j.jacadv.2024.101314] [Medline: 39435181]
21. Lange DC, Conte S, Pappas-Block E, et al. Cancellation of the cardiac catheterization lab after activation for ST-segment-elevation myocardial infarction. Circ Cardiovasc Qual Outcomes 2018 Aug;11(8):e004464. [doi: 10.1161/CIRCOUTCOMES.117.004464] [Medline: 30354373]

Abbreviations

ACS: acute coronary syndrome
CCL: cardiac catheterization laboratory
ECG: electrocardiogram
FMC: first medical contact
HIPAA: Health Insurance Portability and Accountability Act
MUHC: McGill University Health Centre
NSTEMI: non–ST-segment elevation myocardial infarction
STEMI: ST-segment elevation myocardial infarction
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
UTM: unnecessary team mobilization

Edited by A Coristine; submitted 08.May.2025; peer-reviewed by KH Lin, Y Chu; revised version received 03.Apr.2026; accepted 03.Apr.2026; published 19.Jun.2026.

Please cite as:
Gibson W, Al Kindi D, Brouillette F, Aldajani A, Chaabo O, Lachance Y, Akl E, Dandona KB, Martucci G, Pelletier JP, Piazza N, Levett JY, Moran T, Spaziano M
Impact of a Cloud-Based Care Coordination Platform on False Cardiac Catheterization Laboratory Activations and Unnecessary Team Mobilization: Retrospective Cohort Study
JMIR Cardio 2026;10:e76932
URL: https://cardio.jmir.org/2026/1/e76932
doi:10.2196/76932

© William Gibson, Dawoud Al Kindi, François Brouillette, Ahmed Aldajani, Omar Chaabo, Yasmine Lachance, Elie Akl, Kshitij Badal Dandona, Giuseppe Martucci, Jean-Philippe Pelletier, Nicolo Piazza, Jeremy Y Levett, Tomer Moran, Marco Spaziano. Originally published in JMIR Cardio (https://cardio.jmir.org), 19.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited.

XSL•FO
RenderX

The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Multilingual Video Education for Hospitalized Patients With Myocardial Infarction (EDUCATE-MI): Single-Arm Implementation Study

Aileen Zeng[1], MScMed; Edel O'Hagan[1], PhD; Sul Ki Kim[2,3], BSc, MBBS; Simone Marschner[1], PhD; Mitchell Sarkies[1,4,5], PhD; Marina Wassif[2], MD; Meng Ji[6], PhD; Julie Ayre[7], PhD; Daniel McIntyre[1], MD, PhD; Clara K Chow[1,2], MBBS, PhD; Aravinda Thiagalingam[1,2*], MBBS, PhD; Liliana Laranjo[1*], MPH, MD, PhD

[1]Westmead Applied Research Centre, Sydney Medical School, The University of Sydney, Level 5, Block K, Westmead Hospital, Hawkesbury Road, Westmead, Australia

[2]Department of Cardiology, Westmead Hospital, Westmead, Australia

[3]Westmead Institute for Medical Research, University of Sydney, Westmead, Australia

[4]School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[5]Sydney Health Partners, Implementation Science Academy, University of Sydney, Sydney, Australia

[6]School of Languages and Cultures, University of Sydney, Sydney, Australia

[7]Sydney Health Literacy Lab, School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Liliana Laranjo, MPH, MD, PhD
Westmead Applied Research Centre, Sydney Medical School, The University of Sydney, Level 5, Block K, Westmead Hospital, Hawkesbury Road, Westmead, Australia

Abstract

Background: Clinical guidelines recommend the early initiation of secondary prevention strategies prior to hospital discharge for patients with myocardial infarction (MI) to reduce morbidity and mortality, but implementation is resource-intensive. Multilingual videos can deliver information in diverse preferred languages and literacy levels, but their impact on MI knowledge among hospitalized patients remains unclear.

Objective: This study aims to assess whether the delivery of a multilingual educational video to hospitalized patients with MI can improve patient MI knowledge before hospital discharge.

Methods: We conducted a single-arm pre-post study with embedded formative implementation evaluation from December 2023 to October 2024 in a tertiary hospital. The intervention was a video on post-MI management, available in English, Arabic, Hindi, and Mandarin (with Simplified Chinese subtitles). The intervention was delivered via a tablet provided by the research assistant. The primary outcome was the change in patient knowledge of MI, measured by comparing the mean number of correct responses before and after the intervention using a 2-tailed paired *t* test. We assessed early-stage implementation using 2 prespecified elements from the Proctor implementation outcomes framework: acceptability and fidelity of the video delivery. We performed content analysis on the notes taken from participants' feedback to improve the video.

Results: We recruited 129 participants (mean age of 59.4, SD 12.6 years) for this study. English was the preferred language (n=96, 74.4%) and Hindi was the predominant non-English language (n=17, 13.2%). Of the 129 participants enrolled, 128 completed follow-up immediately postintervention (1 lost interest). The average number of correct responses out of 10 was 5.4 (SD 2.7) at baseline and 7.2 (SD 2.5) postintervention (mean difference=1.9, 95% CI 1.6-2.2; P<.001; Cohen d_{rm} for paired change=0.72). The educational video was well-accepted, with 83.6% (107/128) of participants finding it easy to understand, 74.2% (95/128) engaging, and 87.5% (112/128) useful. Participants' feedback for improvement highlighted content complexity and a preference for conversational language and dialects. Fidelity of the intervention was subjectively assessed as reasonably achieved, given that the core components of the intervention (ie, animations and educational content conveyed through the audio and subtitles) were delivered as intended. Fidelity of the implementation strategy was similarly assessed as reasonably achieved because there were no technology issues preventing delivery of the intervention as intended, through video display from a weblink embedded in REDCap, using a tablet with internet connection.

Conclusions: A short educational video may improve patient knowledge of MI before discharge. Further scaled research is needed to evaluate the effectiveness and implementation of this intervention in additional languages and diverse populations.

XSL•FO
RenderX

This study highlights the need for culturally and linguistically tailored resources in clinical settings, informing future research and policy on inclusive patient education.

(JMIR Cardio 2026;10:e82817) doi:10.2196/82817

KEYWORDS

myocardial infarction; health education; diversity, equity, inclusion; DEI; multilingual video

Introduction

Cardiovascular disease (CVD), a group of diseases affecting the heart and blood vessels, remains the leading cause of death worldwide [1]. Most CVD mortality is due to atherosclerotic causes, including acute myocardial infarction (MI) secondary to plaque rupture (type 1 MI), accounting for 38% to 44% of all CVD-related deaths worldwide [2]. After the acute phase, the risk of death from another cardiac event remains high [3].

International guidelines recommend the early initiation of secondary prevention strategies prior to hospital discharge to reduce the risk of morbidity and mortality [4], but there are challenges in implementation. A key component of secondary prevention is adherence to various medications, including long-term aspirin and dual antiplatelet therapy [5]. However, adherence to these medications remains suboptimal [6] due to barriers including psychological factors and the complexity of the drug regimen [7]. Patient education is key to addressing MI knowledge gaps and suboptimal medication adherence, but inpatient education delivery relies on time-pressured frontline clinicians and may fail to accommodate the needs of diverse populations [8].

Educational videos can help deliver evidence-based health information to diverse populations and accommodate different language preferences and literacy levels, including people with limited reading ability or those who are illiterate [9,10]. A systematic review (59 experimental studies, n=9789) of video-based educational interventions reported improved knowledge in 75 % (30/40) of the assessed outcomes in people with chronic diseases [9]. Previous studies evaluating inpatient video interventions for patients post-MI [11,12] have shown improvements in patient knowledge, but the videos in these studies were delivered in 1 predominant language. To date, no studies have evaluated the impact of multilingual educational videos delivered during admission on patient knowledge post-MI. Assessing knowledge during hospitalization and before discharge is important as distress in the acute period of an MI can interfere with understanding and recalling clinical information [13].

The aim of this study was to assess whether the delivery of a multilingual educational video to hospitalized patients with MI can improve their patient knowledge of MI before hospital discharge.

Methods

Study Design

We conducted a single-arm pre-post study with embedded formative implementation evaluation in a tertiary teaching hospital in Sydney, Australia. The protocol is available on Open Science Framework (registered on April 12, 2024) [14]. Reporting of this study follows the StaRI (Standards for Reporting Implementation Studies; Checklist 1), TIDieR (Template for Intervention Description and Replication; Checklist 2) guidelines, and TREND (Transparent Reporting of Evaluations with Non-Randomized Designs; Checklist 3) [15-17].

Patient and Public Involvement

Members of the public were not involved in the design of the study or the interpretation of the findings. There are plans to disseminate the results of the research to study participants and the community via our institute's monthly newsletter.

Participants and Setting

Patients were invited to participate if they were aged 18 years or older, had been admitted for inpatient services at Westmead Hospital for a type 1 MI [18], had undergone coronary angiography, and understood 1 of the 4 available languages: English, Arabic, Hindi, or Mandarin (with Simplified Chinese subtitles; Multimedia Appendix 1).

Participants who were unable to consent to the study in 1 of 4 four languages or who were unable to complete the video due to cognitive or visual impairments were excluded.

Intervention and Implementation Strategy

The intervention was a single video with subtitles, approximately 5 minutes long, with voiceovers performed by native speakers in English, Hindi, Arabic, and Mandarin (Simplified Chinese subtitles) to ensure linguistic appropriateness. The video included different animations and focused on explaining the disease process leading to a type 1 MI (ie, acute coronary atherothrombosis) and the importance of different medications post-MI. The video was developed by a consultant cardiologist and a cardiology advanced trainee based on the latest clinical guidelines, using whiteboard animation software, with input from a multidisciplinary team. The development and description of the video is described following the TIDieR checklist in [16].

The implementation strategy was the delivery of the intervention via a tablet with an internet connection, provided by the research assistant. The video was embedded in REDCap (Research Electronic Data Capture; Vanderbilt University) surveys (accessed via a weblink) and hosted online.

Recruitment and Data Collection

Eligible participants were identified through discussion with their care team and approached at the patient's bedside. Participants were informed that they could withdraw at any time. We did not approach participants who were clinically unstable, as indicated by the ward clinicians. A research assistant

XSL•FO
RenderX

conducted the consent process and stayed with the patient while they watched the video. Patients were encouraged to provide feedback on the video during and after visualization; the research assistant took notes of patients' comments and feedback. At baseline, sociodemographic data and MI knowledge were collected. MI knowledge was assessed immediately postintervention and at a 1-month follow-up (the latter being optional, for participants who opted to provide an email address for this purpose), as well as data on the acceptability of the video. Sociodemographic data, MI knowledge, and acceptability data were collected via self-reported electronic questionnaires hosted on the REDCap platform.

Study Outcomes

The prespecified primary outcome was prospectively defined as the average number of correct responses on the MI knowledge questionnaire immediately after the intervention, compared to baseline. Given the lack of validated questionnaires assessing MI knowledge [19], one of the investigators (AT) developed the MI knowledge questionnaire with contributions from other clinicians. The tool is a 10-item multiple-choice (5 options with 1 correct answer per question) questionnaire that assesses general MI knowledge, with questions on the causes of MI and post-MI medications (with the total number of correct responses ranging from 0="no correct answers" to 10=" all answers to the 10 questions were correct"; Multimedia Appendix 1).

A prespecified secondary outcome was the average number of correct responses in the MI knowledge questionnaire at 1 month. Exploratory post hoc outcomes included the proportion of participants who improved their total number of correct responses from baseline to postintervention, the proportion meeting the knowledge target (defined as 7 or more correct responses out of 10) or medication knowledge target (defined as correctly answering all 7 medication questions), and the proportion of correct responses for each of the 10 individual questions postintervention compared to baseline.

To assess early-stage implementation, we evaluated two prespecified elements from the Proctor implementation outcomes framework [20]: (1) acceptability of the intervention and (2) fidelity of the intervention and implementation strategy (ie, delivery of the intervention via a tablet with an internet connection). Acceptability of the intervention was assessed through content analysis of participants' feedback (described in the "Data Analysis" section) and via a questionnaire asking participants to rate 3 different statements on a 5-point scale from strongly disagree to strongly agree ("The information delivered in the video was easy enough to understand," "I found the video engaging," and "I found the information useful") and asking whether they would be interested in receiving similar videos in the future (yes or no).

Fidelity of the intervention was subjectively assessed by the investigators based on whether the core components of the intervention (ie, video animations and educational content conveyed through the audio and subtitles) were delivered as intended across the 4 different language versions. For this assessment, perspectives from the translators involved in the study were sought regarding the extent to which the translated content matched the English version. Fidelity of the implementation strategy was subjectively assessed by the investigators based on deviations from the intended mode of delivery (ie, video display from a weblink embedded in REDCap, using a tablet with internet connection).

Data Analysis

Data analysis followed a statistical analysis plan developed a priori (Open Science Framework) [14]. A sample size of 119 participants was estimated to provide 90% power to detect a moderate effect size (d=0.3) in knowledge difference pre-post, considering a 2-sided type 1 error of 0.05.

We expressed descriptive data as proportions and means with SD. The prespecified primary outcome was analyzed using a 2-tailed paired *t* test, measuring the mean difference in patient knowledge before and after the intervention, reported with 95% CIs, and Cohen d_{rm} for paired change (Multimedia Appendix 1). An exploratory McNemar test was used to assess changes in the proportion of participants who met knowledge targets before and immediately postintervention, as well as the proportion of correct responses for each of the 10 individual questions postintervention compared to baseline. A prespecified 2-tailed paired *t* test was used to measure the mean difference in patient knowledge before and 1 month postintervention. To test the homogeneity of the treatment effect across subgroups, we used ANCOVA (postadjusted for baseline) to assess the change in the average number of correct responses by age (≤65 y and >65 y), sex (male and female), language (English and non-English), and education (prior to secondary education completion and postsecondary education; Multimedia Appendix 1). We report the proportion of participants who improved their total number of correct responses from baseline to postintervention and implementation measures of acceptability descriptively immediately postintervention and at 1-month follow-up.

Analyses were conducted using R statistical software (version 4.4.0; R Project for Statistical Computing). *P* value was 2-sided, and statistical significance was set at *P*=.05. Data were analyzed from October 2024 to November 2024. Participants with missing data at follow-up were excluded from the analysis. Two researchers performed content analysis on the notes taken from participants' feedback. We were interested in understanding the participants' perspectives on the content of the video in the different languages and their suggestions for improving the videos. Themes were reviewed and discussed with 2 other investigators to clarify, explore, and refine interpretations.

Ethical Considerations

Ethical approval (2021_ETH00983_v3) was obtained from the Western Sydney Local Health District Human Research Ethics Committee. Participant information and consent forms were available in English, Arabic, Hindi, and Mandarin (Simplified Chinese subtitles). Informed consent was obtained in the participants' preferred language. All documents and content in the intervention were translated by the multicultural unit in the Western Sydney Local Health District by a qualified health interpreter. Data were collected and stored on secure servers accessible only to approved study personnel. All data were

XSL•FO
RenderX

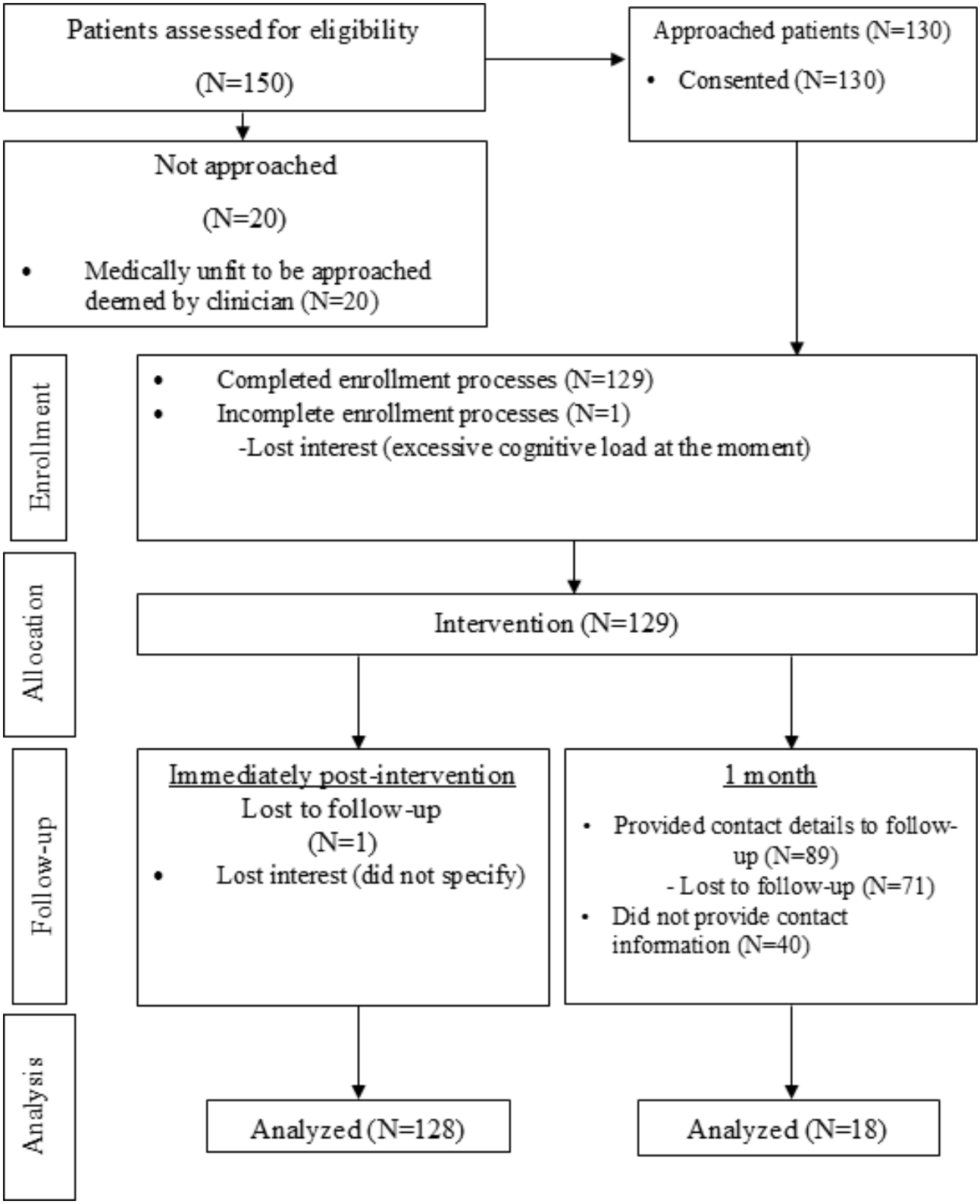
deidentified for data analysis and publication. No compensation was offered to the participants.

Results

Between December 2023 and October 2024, a total of 150 patients were assessed for eligibility. Of these, 130 were approached (20 were deemed by a clinician as not medically fit to be approached for the study), and 129 were enrolled (Figure 1). Of the 129 participants enrolled, 128 completed follow-up immediately postintervention (1 lost interest). Of the 89 participants who provided their email address for the 1-month survey, 18 responded (Figure 1). The baseline characteristics of participants are summarized in Table 1. The mean participant age was 59.4 (SD 12.6) years and 20.2% (26/129) were female. In our sample, 24.8% (32/129) were South Asian and 76% (98/129) completed secondary education or above. The majority of our participants listed English as their preferred language (96/129, 74.4%), followed by Hindi (17/129, 13.2%), Arabic (10/129, 7.8%), and Mandarin (with Simplified Chinese subtitles; 6/129, 4.7%).

XSL•FO
RenderX

Figure 1. Flow diagram of participants in the single-arm pre-post study.



XSL•FO
RenderX

Table . Baseline sociodemographic and clinical characteristics of participants (N=129).

Characteristics[a]	Value
Sex, n (%)	
Female	26 (20.2)
Age (y; missing=3), mean (SD)	59.4 (12.6)
Ethnicity, n (%)	
Aboriginal or Torres Strait Islander	3 (2.3)
Australian or New Zealander	25 (19.4)
Polynesian	1 (0.8)
European	29 (22.5)
American (North, Central, and South)	1 (0.8)
South Asian (Bangladesh, India, Nepal, Pakistan, and Sri Lanka)	32 (24.8)
East Asian (China, Japan, and Taiwan)	8 (6.2)
South-East Asian (Vietnam, Cambodia, Laos, Burma, Malaysia, Singapore, Philippines, Thailand, Indonesia, and East Timor)	8 (6.2)
Middle East and North African	19 (14.7)
Sub-Saharan Africa	1 (0.8)
Pacific Islander	2 (1.6)
Other	0 (0)
Preferred languages, n (%)	
English	96 (74.4)
Arabic	10 (7.8)
Hindi	17 (13.2)
Mandarin (with Simplified Chinese subtitles)	6 (4.7)
Education level, n (%)	
Never attended school	1 (0.8)
Primary	7 (5.4)
Secondary school without completion certificate	23 (17.8)
Secondary school graduate	25 (19.4)
Technical or vocational qualifications	10 (7.8)
University undergraduate	47 (36.4)
University postgraduate	16 (12.4)
Risk factors, n (%)	
Diabetes	48 (37.2)
Hypertension	74 (57.4)
High cholesterol	66 (51.2)
Smoker (or recently quit <12 mo)	47 (36.4)

[a]All data are self-reported.

In the primary outcome analysis, the average number of correct responses was 5.4 (SD 2.7) at baseline and 7.2 (SD 2.5) postintervention (mean difference=1.9, 95% CI 1.6-2.2; *P*<.001; Cohen d_{rm} for paired change=0.72; Table 2). At 1 month postintervention, there were 18 completed responses (Multimedia Appendix 1). Overall, 72.7% (93/128) of the participants showed higher scores in their total number of correct responses postintervention compared to baseline (Multimedia Appendix 1). The proportion of participants who met the MI knowledge target postintervention increased from 47/129 (36.4%) to 93/128 (72.7%; *P*<.001; Figure 2). For each of the individual questions, the proportion of correct responses increased from baseline to postintervention, except for 1 question (question 7; Multimedia Appendix 1).

XSL•FO
RenderX

Regarding the acceptability of the intervention, most participants reported that they agreed or strongly agreed that the information delivered in the video was easy to understand (107/128, 83.6%), engaging (95/128, 74.2%), and useful (112/128, 87.5%); additionally 84.4% (108/128) reported they would like to receive similar videos in the future (Multimedia Appendix 1).

Table . Mean difference in myocardial infarction (MI) knowledge before and immediately after the multilingual video intervention in hospitalized patients with MI prior to discharge (n=128).

	Baseline, n=129	Immediately postintervention, n=128	Mean difference, n=128[a] (95% CI)	*P* value	Cohen d_{rm}[b]
Mean number of correct responses[c], mean (SD)	5.4 (2.7)	7.2 (2.5)	1.9 (1.6-2.2)	<.001	0.72

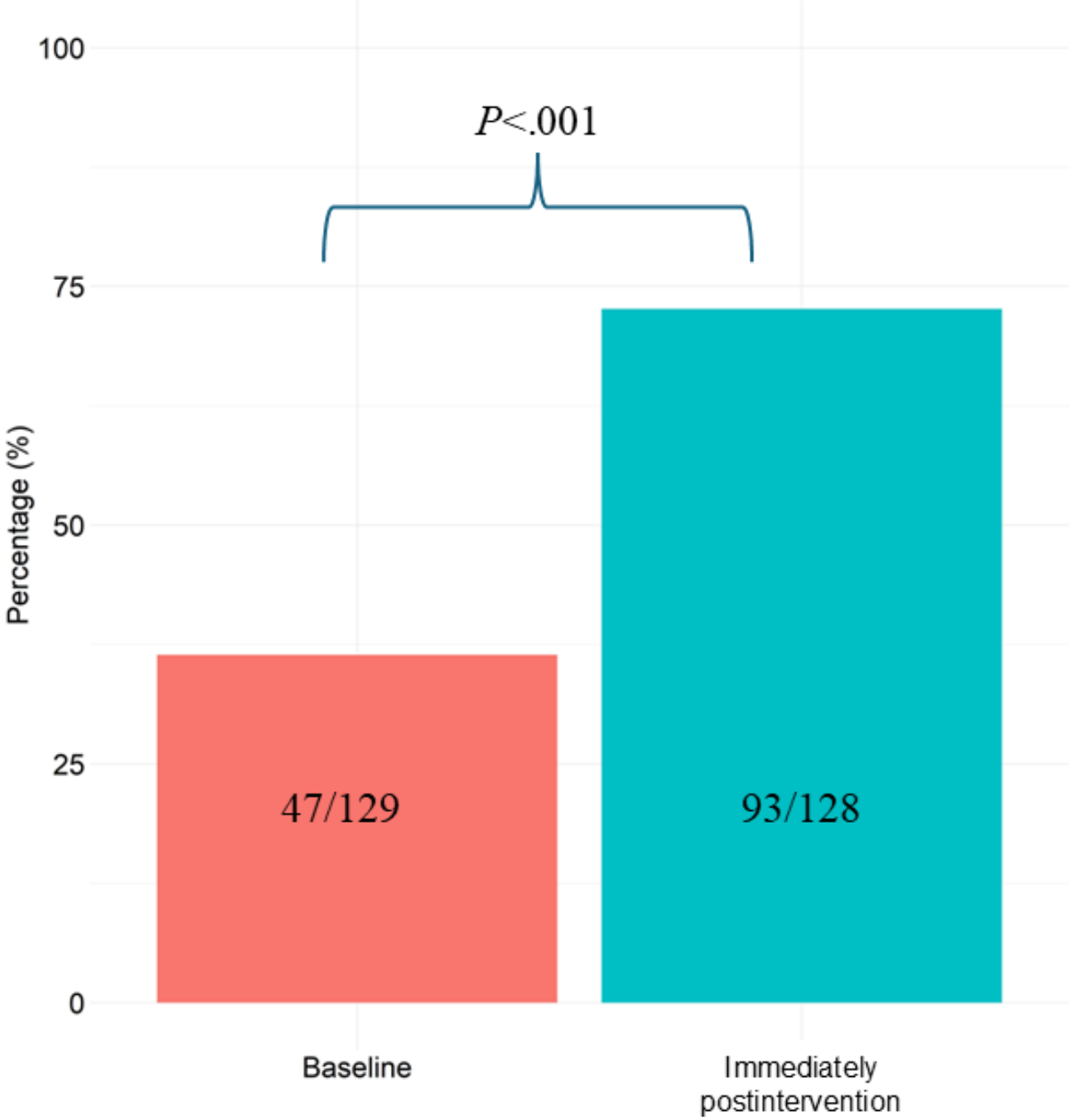
[a]Analyzed using 2-tailed paired *t* test.

[b]Cohen d_{rm} for paired change - adjusted d_z estimating the "standard" between-subjects *d* by a factor of 2(1−r), where *r* is the Pearson correlation between the paired measures [21] (Supplement F in Multimedia Appendix 1).

[c]Correct responses range from 0 to 10.

XSL•FO
RenderX

Figure 2. Bar plot illustrating the proportion of participants who met the knowledge target (defined as 7 or more correct responses out of 10) before and immediately after the video intervention. Changes in the proportion of participants were assessed using a McNemar test.



Nine participants (5 Hindi-speaking, 2 Mandarin-speaking, and 2 Arabic-speaking) provided qualitative feedback on the videos. Content analysis of the feedback revealed two main themes: (1) complexity of the content and medical jargon, and (2) preference for conversational language and dialects. Regarding the complexity of the content and medical jargon, participants indicated they had difficulty understanding some terms in the videos, expressing a preference for lay language. Examples of quotes from participants grouped in this theme include:

> *The video will be too difficult to understand (for people not) familiar with the terminology.* [Study ID 125; 70‐75 years; preferred language: Hindi; educational level: university undergraduate]

> *High-order Hindi was used—needs to have less jargon.* [Study ID 30; 66‐70 years; preferred language: Hindi; educational level: university postgraduate]

Regarding the preference for conversational language and dialects, participants with Hindi as their preferred language reported that a macaronic-hybrid language was commonly used to communicate and would be preferred for the video. Participants with Arabic as their preferred language mentioned that they would prefer the video to be in their specific dialect. Examples of quotes are presented as follows:

> *Mix of Hindi and English would have been better—conversational Hindi.* [Study ID 30; 66-70

XSL•FO
RenderX

years; preferred language: Hindi; educational level: university postgraduate]

Dialect is like Egyptian… parents would understand Nahw… (I) am bilingual—so can't read or write. [Study ID 108; 40-45 years; preferred language: Arabic; educational level: university undergraduate]

Fidelity of the intervention was subjectively assessed as reasonably achieved, given that the core components of the intervention (ie, animations and educational content conveyed through the audio and subtitles) were delivered as intended. The video animations were the same across the 4 translated versions of the educational video; only the timing between different animations changed to accommodate the audio. The educational content conveyed through the audio and subtitles was also delivered as intended, with minor differences in language between the different translations. Fidelity of the implementation strategy was similarly assessed as reasonably achieved because there were no technology issues preventing delivery of the intervention as intended through video display from the weblink embedded in REDCap, using a tablet with internet connection.

Discussion

Principal Findings

In this single-arm pre-post study with embedded formative implementation evaluation of 128 inpatients admitted for an MI, we found that a short multilingual educational video may improve patient knowledge of MI before discharge. The videos were deemed acceptable, with most participants finding them easy to understand, engaging, and useful. The core components of the intervention were delivered as intended, as the videos across the 4 languages comprised the same animations and educational content in both audio and subtitles, with only slight timing differences due to language nuances between translations. There were no deviations from the implementation strategy, with all participants watching the video through a REDCap weblink on a tablet with internet connection, as intended. Our study demonstrates that it is possible to deliver videos in different languages—English, Arabic, Hindi, and Mandarin (with Simplified Chinese subtitles)—to inpatients in a linguistically diverse context at a tertiary hospital, with modest improvement in MI knowledge before discharge.

Comparisons With Prior Work

Results from our study suggest that a single multilingual video intervention may improve patient MI knowledge in the inpatient setting. Our results are consistent with other studies evaluating inpatient video interventions for patients post-MI [11,12], although ours was the only one delivered in more than 1 language. One of these studies, a randomized controlled trial (RCT) of 68 inpatients, reported a moderate within-subject change in the intervention group (15-min video) at 3-month follow-up in the same order of magnitude that we found in our study [11]. A 2-arm quasi-experimental study (N=25) reported a 50% larger within-subject change in the intervention group at 7-day follow-up [12]. These medium to large within-subject changes also appear to persist when video interventions are delivered in a postdischarge cardiac rehabilitation setting. A pre-post study of a culturally adapted secondary prevention video education program (Simplified Chinese) for patients post-MI reported that one-quarter of patients found the information overwhelming, despite a medium within-subject change in knowledge similar in magnitude to our findings [22]. Given the competing demands during MI recovery, even the modest knowledge gains noted may be clinically meaningful and are consistent with previous findings. It remains unclear how different health literacy strategies applied to multilingual videos can influence knowledge in different populations, which should be explored in future research [23].

The mean postintervention score was 7/10, even after participants had just viewed the video, highlighting the persistent challenges of meeting patients' educational and health literacy needs. Despite efforts to adhere to readability recommendations (ie, grade 8 or below), our findings indicate additional attention is required to produce simpler content with fewer unfamiliar terms. These findings were echoed in our feedback discussions and may reflect broader barriers to comprehending health resources, such as high readability levels (ie, above the recommended grade 8 level) and the use of medical jargon [24]. Reducing readability levels prior to translation into non-English languages is important, as content complexity may be compounded during translation [24,25]. In addition, the literal translation of certain terms may not accurately convey their intended meaning, and the use of culturally equivalent terms may be preferable [25,26]. Hence, cultural adaptation is key to adequately considering the cultural aspects that may not be captured through linguistic translation alone, such as cultural equivalence, cultural appropriateness, similarity of interpretability, and item relevance [25,27].

While knowledge scores improved postintervention and the video in our study was well-accepted, further gains may be achieved with the involvement of community members in the co-design and cultural adaptation of intervention content [24,28]. Notwithstanding the importance of addressing language barriers in improving health outcomes and quality of care [29], future studies aiming to develop equitable health education interventions should also consider the nuances of social context and how language intersects with race, migration status, religion, socioeconomic position, and other social determinants of health [30-32].

Strengths and Limitations

Strengths of this study include delivering the intervention to a diverse, multicultural population (n=114, 88% non-Caucasian) in an inpatient setting. We translated the intervention into 4 different languages and adapted the medical jargon in the translations. However, there were challenges in translating medical language, and the readability of our English-language content was high, at grade 10, instead of the desired grade 8. These limitations are common in medical content, where the inclusion of specific, complex, condition-specific terminology (eg, MI) is often unavoidable and increases overall readability scores [33].

This study should be interpreted in the context of its single-arm pre-post design, which limits the evaluation of causality. Our English educational content did not meet the recommended grade 8 readability levels and was at a grade 10 level due to the

XSL•FO
RenderX

inclusion of unavoidable medical terms, which were otherwise explained in the transcript. Translations were reviewed by native speakers, but more comprehensive patient involvement in co-design and cultural adaptation will be key for future iterations, based on participant feedback. Only 34 out of 129 participants received the intervention in a language other than English, which reflects the known challenges of recruiting participants from diverse backgrounds [34,35]. Furthermore, our content analysis of patient feedback was based on researcher notes rather than verbatim transcripts of recorded feedback, limiting opportunities for in-depth analysis of participants' perspectives of the intervention. Patient knowledge may have been influenced by short-term recall in the immediate postintervention period, and the 1-month results should be interpreted with caution due to low response rates and the risk of survivor bias. Given the absence of validated patient questionnaires to assess MI knowledge, the MI knowledge questionnaire was created by clinicians in the study and was not psychometrically assessed. It may not measure actionable MI knowledge, which is key for patient empowerment and behavior change. Finally, we did not assess medication adherence, secondary prevention behaviors, health literacy, or language proficiency, which could have aided in the interpretation of the results.

Multilingual videos have the potential to reduce language barriers in the delivery of inpatient education for patients with MI prior to discharge, particularly in health care services that serve multicultural communities. We demonstrated that it is possible to improve knowledge in the short term and implement educational videos within a 5-minute timeframe for hospitalized patients with MI in a multicultural context. This offers a scalable and pragmatic strategy that may ease demands on health care staff prior to discharge. Patients report a considerable treatment burden in the first year following an MI, particularly those with lower health literacy [36], which may contribute to suboptimal medication adherence. Providing education prior to discharge may help alleviate this burden and serve as a primer for adherence to secondary prevention care. In the future, large language models and other generative artificial intelligence tools could complement video-based education by providing interactive, personalized support [37] in a patient's preferred language and at an appropriate readability level [38]. These tools could enhance health literacy and support equitable care for patients from diverse linguistic backgrounds. Future research, particularly larger RCTs with longer follow-up, should consider strategies to better engage participants with non-English language preferences, to enhance inclusivity, and evaluate the effectiveness of multilingual patient education interventions in different language groups.

Conclusions

A short educational video delivered during inpatient hospital admission for acute MI may improve patient knowledge before discharge. Our study demonstrates that it is possible to deliver videos in different languages—English, Arabic, Hindi, and Mandarin (with Simplified Chinese subtitles)—to patients in a multicultural context at a tertiary hospital. Future interventions should be culturally adapted and co-designed with patients. Further RCTs are needed to evaluate the long-term impact of this intervention on MI knowledge and secondary prevention behaviors across different settings.

Acknowledgments

The authors acknowledge research officer Wendy Liu for her contribution in providing the Mandarin voiceover for one of the videos and medical officer Dr. Rukmini Kulkarni for her contribution in providing the Hindi voiceover. The authors also acknowledge the nursing staff and doctors from the Westmead Cardiology Unit for facilitating recruitment efforts, as well as the biostatistical team at the Westmead Applied Research Centre (Haeri Min and Dr. Desi Quintans) for validating the results.

Funding

This work was supported by a Sydney Health Partners Implementation Science Pilot Grant 2021 (grant approval: clinician created multimedia and multicultural cardiovascular m-Health education: EDUCATE_MI). LL is supported by a National Health and Medical Research Council (NHMRC) Investigator Grant (grant 2017642) and Sydney Horizon Fellowship. MS is supported by an NHMRC Investigator Grant (grant 2007970) and Sydney Horizon Fellowship. CKC is supported by an NHMRC Investigator Grant (grant APP1195326). JA is supported by an NHMRC Investigator Grant (grant 2017278). SKK is supported by a Postgraduate Research Scholarship in Stem Cells and Regenerative Medicine and Western Sydney Local Health District JMO Research Scholarship.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AZ was responsible for data collection, data analysis, and writing the manuscript. AT, LL, and SKK contributed to the conceptualization and study design and critically reviewed the manuscript. EO assisted in the interpretation of the data analysis and critically reviewed the manuscript. SM assisted in the data analysis and critically reviewed the manuscript. MS, MW, JA, DM, and CKC critically reviewed and contributed to the manuscript. AT and LL are co-senior authors.

XSL•FO
RenderX

Conflicts of Interest

JA is a co-director of Health Literacy Solutions Pty Ltd. The SHeLL Editor is a research tool owned by the University of Sydney and is sublicensed to Health Literacy Solutions Pty Ltd to support broader public access and use.
CKC is a recipient of an Investigator Grant from the National Health and Medical Research Council and is an investigator on current grants funded by the Medical Research Future Fund and the National Health and Medical Research Council. CKC is a member of the Board of the Western Sydney Local Health District and the National Heart Foundation of Australia. CKC has previously received research funding or grants from NSW Health, the Australian Digital Health Agency, and Google. CKC has also previously received speaker fees from several organizations, including Novartis, Limbic, Eli Lilly and Company, Novo Nordisk, and Amgen. The George Institute for Global Health has submitted patent applications related to low fixed-dose combination products for the treatment of cardiovascular and cardiometabolic diseases. CKC is listed as one of the inventors on these applications. She reports no direct financial interests in the patent applications or related investments.
The other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary material (Setting Details, Protocol Deviations, MI Knowledge Questionnaire, and Detailed Analyses).
[DOCX File, 94 KB - cardio_v10i1e82817_app1.docx]

Checklist 1
StaRI checklist.
[DOCX File, 55 KB - cardio_v10i1e82817_app2.docx]

Checklist 2
TiDier checklist.
[DOCX File, 276 KB - cardio_v10i1e82817_app3.docx]

Checklist 3
TREND checklist.
[DOCX File, 53 KB - cardio_v10i1e82817_app4.docx]

Checklist 4
iCHECK-DH checklist.
[DOCX File, 28 KB - cardio_v10i1e82817_app5.docx]

References

1. Murray CJL. The Global Burden of Disease Study at 30 years. Nat Med 2022 Oct;28(10):2019-2026. [doi: 10.1038/s41591-022-01990-1] [Medline: 36216939]
2. Timmis A, Vardas P, Townsend N, et al. European Society of Cardiology: cardiovascular disease statistics 2021. Eur Heart J 2022 Feb 22;43(8):716-799. [doi: 10.1093/eurheartj/ehab892] [Medline: 35016208]
3. Nair R, Johnson M, Kravitz K, et al. Characteristics and outcomes of early recurrent myocardial infarction after acute myocardial infarction. J Am Heart Assoc 2021 Aug 17;10(16):e019270. [doi: 10.1161/JAHA.120.019270] [Medline: 34333986]
4. Byrne RA, Rossello X, Coughlan JJ, et al. ESC guidelines for the management of acute coronary syndromes: developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). Eur Heart J 2023;44(38):3720-3826. [doi: 10.1093/eurheartj/ehad191] [Medline: 37622654]
5. Laranjo L, Lanas F, Sun MC, et al. World Heart Federation roadmap for secondary prevention of cardiovascular disease: 2023 update. Glob Heart 2024;19(1):8. [doi: 10.5334/gh.1278] [Medline: 38273995]
6. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-887. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]
7. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314. [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
8. Geerligs L, Rankin NM, Shepherd HL, Butow P. Hospital-based interventions: a systematic review of staff-reported barriers and facilitators to implementation processes. Implement Sci 2018 Dec;13(1):36. [doi: 10.1186/s13012-018-0726-9]
9. Deshpande N, Wu M, Kelly C, et al. Video-based educational interventions for patients with chronic illnesses: systematic review. J Med Internet Res 2023 Jul 19;25:e41092. [doi: 10.2196/41092] [Medline: 37467015]
10. Sarkies MN, Maloney S, Symmons M, Haines TP. Video strategies improved health professional knowledge across different contexts: a helix counterbalanced randomized controlled study. J Clin Epidemiol 2019 Aug;112:1-11. [doi: 10.1016/j.jclinepi.2019.04.003] [Medline: 30998965]

XSL•FO
RenderX

11. Bruggmann C, Adjedj J, Sardy S, Muller O, Voirol P, Sadeghipour F. Effects of the interactive web-based video “Mon Coeur, Mon BASIC” on drug adherence of patients with myocardial infarction: randomized controlled trial. J Med Internet Res 2021 Aug 30;23(8):e21938. [doi: 10.2196/21938] [Medline: 34459744]
12. Winiger AM, Shue-McGuffin K, Moore-Gibbs A, Jordan K, Blanchard A. Implementation of an Ask Me 3 ® education video to improve outcomes in post-myocardial infarction patients. Am J Prev Cardiol 2021 Dec;8:100253. [doi: 10.1016/j.ajpc.2021.100253] [Medline: 34568857]
13. Whitehead DL, Perkins-Porras L, Strike PC, Steptoe A. Post-traumatic stress disorder in patients with cardiac disease: predicting vulnerability from emotional responses during admission for acute coronary syndromes. Heart 2006 Sep;92(9):1225-1229. [doi: 10.1136/hrt.2005.070946] [Medline: 16424065]
14. Clinician-created multimedia and multicultural cardiovascular m-health education—EDUCATE-MI. Open Science Framework. URL: https://osf.io/xf92g/ [accessed 2025-08-22]
15. Pinnock H, Barwick M, Carpenter CR, et al. Standards for Reporting Implementation Studies (StaRI) statement. BMJ 2017 Mar 6;356:i6795. [doi: 10.1136/bmj.i6795] [Medline: 28264797]
16. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
17. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
18. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). Circulation 2018 Nov 13;138(20):e618-e651. [doi: 10.1161/CIR.0000000000000617] [Medline: 30571511]
19. Fajardo MA, Yung C, Cornell S, et al. Quality of available cardiovascular disease knowledge tools: a systematic review. Glob Heart 2025;20(1):59. [doi: 10.5334/gh.1446] [Medline: 40655032]
20. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
21. Ben-Shachar M, Lüdecke D, Makowski D. Effectsize: estimation of effect size indices and standardized parameters. J Open Source Softw 2020;5(56):2815. [doi: 10.21105/joss.02815]
22. Shi W, Zhang L, Ghisi GLM, Panaretto L, Oh P, Gallagher R. Evaluation of a digital patient education programme for Chinese immigrants after a heart attack. Eur J Cardiovasc Nurs 2024 Sep 5;23(6):599-607. [doi: 10.1093/eurjcn/zvad128] [Medline: 38211942]
23. The Patient Education Materials Assessment Tool (PEMAT) and user’s guide. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/health-literacy/patient-education/pemat.html [accessed 2025-06-17]
24. Shan Y, Ji M, Dong Z, Xing Z, Wang D, Cao X. The Chinese version of the Patient Education Materials Assessment Tool for printable materials: translation, adaptation, and validation study. J Med Internet Res 2023 May 18;25(1):e39808. [doi: 10.2196/39808] [Medline: 37200085]
25. Shan Y, Xing Z, Dong Z, Ji M, Wang D, Cao X. Translating and adapting the DISCERN instrument into a simplified Chinese version and validating its reliability: development and usability study. J Med Internet Res 2023 Feb 2;25(1):e40733. [doi: 10.2196/40733] [Medline: 36729573]
26. Thakkar J, Karthikeyan G, Purohit G, et al. Development of macaronic Hindi-English “Hinglish” text message content for a coronary heart disease secondary prevention programme. Heart Asia 2016;8(2):32-38. [doi: 10.1136/heartasia-2016-010789] [Medline: 27752288]
27. Mac O, Ayre J, McCaffery K, Boroumand F, Bell K, Muscat DM. The readability study: a randomised trial of health information written at different grade reading levels. J Gen Intern Med 2025 Jun;40(8):1820-1828. [doi: 10.1007/s11606-024-09200-z] [Medline: 39707098]
28. Henderson S, Kendall E, See L. The effectiveness of culturally appropriate interventions to manage or prevent chronic disease in culturally and linguistically diverse communities: a systematic literature review. Health Soc Care Community 2011 May;19(3):225-249. [doi: 10.1111/j.1365-2524.2010.00972.x] [Medline: 21208326]
29. Diamond L, Izquierdo K, Canfield D, Matsoukas K, Gany F. A systematic review of the impact of patient-physician non-English language concordance on quality of care and outcomes. J Gen Intern Med 2019 Aug;34(8):1591-1606. [doi: 10.1007/s11606-019-04847-5] [Medline: 31147980]
30. Reppas-Rindlisbacher C, Rawal S. Understanding linguistic inequities in healthcare: moving from the technical to the social. BMJ Qual Saf 2024 Jun 19;33(7):416-418. [doi: 10.1136/bmjqs-2023-016873] [Medline: 38609172]
31. Abdi I, Tinessia A, Mahimbo A, Sheel M, Leask J. Is it time to retire the label “CALD” in public health research and practice? Med J Aust 2025 Mar 17;222(5):220-222. [doi: 10.5694/mja2.52608] [Medline: 39953844]
32. Javed Z, Haisum Maqsood M, Yahya T, et al. Race, racism, and cardiovascular health: applying a social determinants of health framework to racial/ethnic disparities in cardiovascular disease. Circ Cardiovasc Qual Outcomes 2022 Jan;15(1):e007917. [doi: 10.1161/CIRCOUTCOMES.121.007917] [Medline: 35041484]

XSL•FO
RenderX

33. Ayre J, Bonner C, Muscat DM, et al. Multiple automated health literacy assessments of written health information: development of the SHeLL (Sydney Health Literacy Lab) Health Literacy Editor v1. JMIR Form Res 2023 Feb 14;7:e40645. [doi: 10.2196/40645] [Medline: 36787164]
34. Watson E, Gulline H, Jane SM, Woollett A, Ayton D. Improving participation of culturally and linguistically diverse participants in clinical trials: an expert consultation. Trials 2025 Mar 25;26(1):105. [doi: 10.1186/s13063-025-08803-z] [Medline: 40133956]
35. Knepper TC, McLeod HL. When will clinical trials finally reflect diversity? Nature 2018 May;557(7704):157-159. [doi: 10.1038/d41586-018-05049-5] [Medline: 29743700]
36. Qu W, Wang X, Zhang S, et al. Factors related to the treatment burden of patients with coronary heart disease: a cross-sectional study. Heart Lung 2025;70:141-146. [doi: 10.1016/j.hrtlng.2024.11.019] [Medline: 39667309]
37. O'Hagan E, McIntyre D, Laranjo L. The potential for a chat-based artificial intelligence model to facilitate educational messaging on hypertension. Hypertension 2023 Aug;80(8):e128-e130. [doi: 10.1161/HYPERTENSIONAHA.123.21395] [Medline: 37325936]
38. Ayre J, Mac O, McCaffery K, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. J Gen Intern Med 2024 Mar;39(4):573-577. [doi: 10.1007/s11606-023-08469-w] [Medline: 37940756]

Abbreviations

CVD: cardiovascular disease
MI: myocardial infarction
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
StaRI: Standards for Reporting Implementation Studies
TIDieR: Template for Intervention Description and Replication
TREND: Transparent Reporting of Evaluations with Non-Randomized Designs

Edited by A Coristine; submitted 22.Aug.2025; peer-reviewed by A Shivanna, E Huriani, F Owoseje; revised version received 17.Feb.2026; accepted 19.Feb.2026; published 26.Mar.2026.

Please cite as:
Zeng A, O'Hagan E, Kim SK, Marschner S, Sarkies M, Wassif M, Ji M, Ayre J, McIntyre D, Chow CK, Thiagalingam A, Laranjo L Multilingual Video Education for Hospitalized Patients With Myocardial Infarction (EDUCATE-MI): Single-Arm Implementation Study
JMIR Cardio 2026;10:e82817
URL: https://cardio.jmir.org/2026/1/e82817
doi:10.2196/82817

© Aileen Zeng, Edel O'Hagan, Sul Ki Kim, Simone Marschner, Mitchell Sarkies, Marina Wassif, Meng Ji, Julie Ayre, Daniel McIntyre, Clara K Chow, Aravinda Thiagalingam, Liliana Laranjo. Originally published in JMIR Cardio (https://cardio.jmir.org), 26.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improving Cardiometabolic Health in Adolescent Girls With Elevated Anxiety: Web-Based Multisite, Pilot, and Feasibility Randomized Controlled Trial

Caitlin Edwards[1,2*], MA, PhD; Lisa Ranzenhofer[3*], PhD; Jason Lavender[4,5*], PhD; Lauren D Gulley[2,6*], PhD; Stephen Aichele[1*], PhD; Isabel Thorstad[4*], BS; Natalia Sanchez[1*], MPH; Ruby Schrag[4*], BS; Zoe Sinkford[4*], BA; Jill Emerick[7*], MD; Victoria Thomas[4*], BS; Thomas Arnold[4*], SCCM; Jami Young[8*], PhD; Denise Wilfley[9*], PhD; Mark Haigney[4*], MD; Lauren Shomaker[1,2*], PhD; Marian Tanofsky-Kraff[4,10*], PhD

[1]Department of Human Development and Family Studies, College of Health and Human Services, Colorado State University, 410 W. Pitkin St, 1570 Campus Delivery, Fort Collins, CO, United States

[10]Department of Medical and Clinical Psychology, Uniformed Services University, Bethesda, MD, United States

[2]Endocrinology, Department of Pediatrics, University of Colorado Anschutz Medical Campus and Children's Hospital Colorado, Aurora, CO, United States

[3]Department of Psychiatry, University at Buffalo Jacobs School of Medicine and Biomedical Sciences, Buffalo, NY, United States

[4]Department of Medicine, Military and Cardiovascular Outcomes Research (MiCOR), Uniformed Services University of the Health Sciences, Bethesda, MD, United States

[5]The Metis Foundation, San Antonio, TX, United States

[6]University of Colorado Anschutz Medical Campus, 6Barbara Davis Center for Diabetes, Aurora, CO, United States

[7]Department of Pediatrics, Uniformed Services University, Bethesda, MD, United States

[8]Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[9]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, United States

[*]all authors contributed equally

Corresponding Author:
Lauren Shomaker, PhD
Department of Human Development and Family Studies, College of Health and Human Services, Colorado State University, 410 W. Pitkin St, 1570 Campus Delivery, Fort Collins, CO, United States

Abstract

Background: Anxiety is common in adolescents with elevated BMI (kg/m^2), affecting 33% - 43% of adolescents with overweight or obesity (age- and sex-specific BMI ≥85th percentile). Anxiety has been associated with poorer cardiometabolic health, and disinhibited eating in response to anxiety may be a key explanatory mechanism. Therefore, reducing anxiety in adolescents may improve disinhibited eating patterns and prevent worsening cardiometabolic health.

Objective: This pilot study primarily aimed to assess the feasibility and acceptability of 2 virtual group psychological interventions, interpersonal psychotherapy (IPT) and cognitive behavioral therapy (CBT), in adolescent girls with elevated anxiety and above-average weight. Secondary aims were to describe changes in anxiety, eating behavior, and cardiometabolic health indices.

Methods: This 2-site, 2-year, pilot and feasibility randomized controlled trial enrolled 12 - to 17-year-old girls with elevated anxiety symptoms (State-Trait Anxiety Inventory for Children total score ≥32) and above-average weight (BMI ≥75th percentile for age and sex), who were in good general health and not concurrently receiving therapy or medication treatment. Girls were randomly assigned to 1 of 2 web-based arms: 12-week, psychologist-led, group IPT or CBT. In both arms, groups of 5 - 7 adolescents met for 90 minutes on secure Zoom (Zoom Communications, Inc) once per week. Anxiety, disinhibited eating, BMI indices, blood pressure, lipids, hemoglobin A_{1c} (HbA_{1c}), and fasting glucose were measured at baseline, 12-week follow-up, and 1-year follow-up. Feasibility of recruitment, retention, and study protocol, and acceptability and fidelity of IPT and CBT were evaluated.

Results: Of 47 eligible adolescents, 40 (85%) enrolled. Enrollment of the total sample (n=40) was completed in 7 months. Retention was 85% (n=17) in IPT and 100% (n=20) in CBT at 12-week follow-up. Protocol adherence exceeded 96% for all intervals. Missing data varied, with higher rates for phlebotomy (up to 30%). Acceptability ratings were above-average for likability (mean 3.03, SD 1.14) and credibility (mean 3.26, SD 0.75; 0 - 4 scale with 4=Highest) across arms. Most adolescents received ≥80% intervention dosage (IPT: 90%, CBT: 95%). Expert fidelity ratings of IPT were 78% (SD 11%) and CBT were 94% (SD 9%). From baseline to 1-year, IPT demonstrated improvements in general anxiety (95% CI −8.95 to −1.40), social

XSL•FO
RenderX

anxiety (95% CI −12.37 to −1.88), and HbA_{1c} (95% CI −0.48 to −0.23); CBT demonstrated improvement in emotional eating (95% CI −1.18 to −0.18), and both arms demonstrated improvements in BMI percentile (effect sizes=−0.82 and −0.63, IPT and CBT, respectively).

Conclusions: This pilot and feasibility trial provided initial support for the feasibility of recruitment and retention and acceptability of web-based group IPT and CBT, as well as yielded areas for optimizing protocol feasibility and intervention fidelity. These preliminary results inform the planning of larger-scale efficacy trials with powered samples to assess effects on BMI and cardiometabolic health indicators among adolescents with elevated anxiety and above-average weight.

Trial Registration: ClinicalTrials.gov NCT05038033; https://clinicaltrials.gov/study/NCT05038033

(JMIR Cardio 2026;10:e82781) doi:10.2196/82781

KEYWORDS

anxiety; adolescents; cardiometabolic health; pilot trial; feasibility; interpersonal psychotherapy; cognitive behavioral therapy; eating behaviors

Introduction

Overweight or Obesity in Adolescence

Elevated BMI (kg/m²) in adolescents is a serious public health concern. In the United States, about one-third of adolescents have overweight or obesity [1], and globally, adolescent obesity has increased by 1.5% annually over the past decade [2]. Overweight or obesity contributes to earlier manifestation and rising prevalence of cardiometabolic diseases such as type 2 diabetes and cardiovascular diseases [3,4]. Adolescents with overweight or obesity are more likely to develop insulin resistance, hypertension, dyslipidemia, and prediabetes [3,4]. Such cardiometabolic health concerns disproportionately impact individuals from historically disadvantaged identities [4,5], and rates are generally higher among girls than among boys [5]. Moreover, Hispanic, Latinx, and Black or African American individuals also experience higher rates of cardiometabolic and cardiovascular health problems across the lifespan, often beginning in adolescence [5,6], and resulting in a significant disease burden [2,6].

Adolescence is a sensitive period for the course of excess weight gain and cardiometabolic health [7], as there are significant changes in nutrition, sleep, and physical activity throughout adolescence [8]. Adolescence is also marked by physical changes in body composition that accompany puberty as well as psychosocial changes, including increases in perceived stress [9] and negative affect (eg, anxiety and depression) [10]. Rates of anxiety and depression are increasing among adolescents, with up to 60% of adolescents reporting symptoms of depression and 30% reporting anxiety [11]. Importantly, adolescents with elevated BMI are at higher risk for anxiety disorders, with 33%-43% of adolescents with overweight or obesity reporting elevated symptoms of anxiety [12-14]. Anxiety symptoms, especially those associated with social anxiety, commonly manifest during adolescence and are more prevalent in girls [15]. The dynamic physical and psychosocial transitions that characterize adolescence present both a period of heightened risk and an opportunity for impactful intervention to prevent excess weight gain and worsening cardiometabolic health by addressing anxiety.

Cardiometabolic Disease Prevention

The typical behavioral approach to cardiometabolic disease prevention is characterized by one-size-fits-all intensive lifestyle treatment focused on decreasing caloric intake and increasing physical activity [16]. However, such approaches have faced challenges with attrition and sustained effectiveness in adolescents, necessitating more targeted and preventative approaches [16]. As an alternative, addressing psychosocial factors underlying eating and other health behaviors important for excess weight gain and cardiometabolic health may be particularly suitable for adolescents with above-average BMI.

Anxiety is posited to affect cardiometabolic health through anxiety-induced behaviors. Anxiety is associated with disinhibited eating (ie, the propensity to overeat in response to stimuli such as appetizing foods or negative emotions) [17-19]. Indeed, emotional dysregulation, such as anxiety, impacts eating decisions via reward circuitry and interoception; palatable foods become more salient during moments of negative affect [20]. Higher anxiety symptoms are also associated with more frequent disinhibited eating [21,22], especially loss-of-control eating [23,24], a disordered eating behavior characterized by a perceived inability to control intake regardless of the amount consumed [24,25]. Furthermore, disinhibited eating patterns, including loss of control and eating in response to negative affect, are associated with greater weight gain, adiposity, and higher insulin resistance in youth [23-26]. Thus, anxiety during adolescence can trigger disinhibited eating leading to excess weight, adiposity, and cardiometabolic health problems.

Frequent anxiety also activates stress-response biobehavioral systems, including the hypothalamic-pituitary-adrenal axis [27], sympathetic-adrenal-medullary system [28], and increased inflammatory response [29]. Repeated and sustained experiences of anxiety contribute to persistent hypothalamic-pituitary-adrenal activation, thereby increasing cortisol [30], which leads to visceral fat accumulation, increased gluconeogenesis, endothelial dysfunction, and dyslipidemia [30]. The ongoing activation of the sympathetic-adrenal-medullary system is associated with increased rates of hypertension, resting heart rate, and decreased heart rate variability [31,32]. Furthermore, chronic anxiety results in upregulated proinflammatory cytokines [33], thereby lowering insulin resistance, increasing atherogenesis, and impairing glucose control [33].

XSL•FO
RenderX

Treatment for anxiety has been shown to normalize stress physiology [34], improve health behaviors [35], and decrease allostatic load [30]. Moreover, psychological interventions to address both mental *and* physical health problems demonstrate greater acceptability than interventions that address mental *or* physical health concerns [36]. Therefore, anxiety represents a distinct treatment target for preventing further excess weight gain and cardiometabolic problems among adolescents with elevated anxiety and above-average BMI, particularly by improving disinhibited eating.

Interventions

Cognitive behavioral therapy (CBT), the standard of care for adolescent anxiety [37,38], has also been shown to reduce disinhibited eating patterns in adolescents [39]. CBT is based on the premise that anxiety is sustained or exacerbated by the interplay of negative thoughts and behavioral avoidance. From this perspective, maladaptive responses to negative affect (eg, disinhibited eating) are learned behaviors that can be replaced with adaptive cognitive and behavioral coping strategies. CBT involves learning to identify and restructure negative thoughts and attitudes as well as using behavioral strategies focused on exposure to anxiety-inducing stimuli and rewards for nonavoidant coping [37,38]. Although CBT is a standard treatment for adolescent anxiety [37,38], its application to improving weight and cardiometabolic health outcomes in adolescents with elevated anxiety and above-average weight has not yet been tested.

Interpersonal psychotherapy (IPT) offers an alternative to CBT. IPT is grounded in interpersonal theory, which holds that social relationships drive negative affect and stress-related behavior [40]. IPT targets 4 interpersonal problem areas, including grief, role transitions, role disputes, and interpersonal difficulties, that sustain or worsen negative affect [40]. Through facilitated and targeted communication exercises, role-plays, and psychoeducation, IPT aims to improve relationships by reducing conflict, enhancing communication, and increasing meaningful relational support, thereby reducing negative affect and stress-related behaviors [40,41]. IPT has been shown to reduce excess weight and adiposity gain over 3 years compared with didactic controls among adolescent girls with elevated anxiety and above-average BMI [42]. Studies in adolescents with disinhibited eating suggest IPT's promise for reducing loss-of-control eating [42], although IPT has yet to be tested with adolescents with anxiety symptoms specifically.

Extant research has compared the efficacy of IPT and CBT when treating adult depression in both randomized clinical trials [43] and in community settings [44], as well as when delivered via telehealth [45]. Generally, IPT and CBT have been found to be equally effective at treating both adult and adolescent anxiety and depression [43,44]. Indeed, IPT and CBT both are considered evidence-based treatments for adolescent mood disorders [45]. Moreover, IPT has been studied in adolescents with disinhibited eating with the aim to prevent excess weight [42,46] and CBT demonstrates clinically significant improvement for adolescents with eating disorders [47].

However, the *interpersonal* focus of IPT and the *intrapersonal* emphasis of CBT have not been specifically compared in adolescents with anxiety and above-average BMI. Both therapies could be expected to decrease anxiety symptoms and, in turn, reduce excess weight gain and associated cardiometabolic health indices by ameliorating disinhibited eating. Yet, to date, no research has been conducted comparing the efficacy of IPT versus CBT for weight and cardiometabolic health, nor are there data elucidating the potential differing mechanisms of change underlying IPT versus CBT.

However, before efficacy or mechanisms can be evaluated, feasibility and acceptability must be established in adolescents with elevated anxiety and above-average BMI. Multisite studies are ideal for maximizing external validity, timely recruitment of a targeted sample, and capacity for long-term follow-up [48]. Mental health interventions delivered via telehealth have been found to be effective in managing diverse mental health conditions among adolescents [49-52]. Additionally, virtually delivered group psychotherapy is a cost-effective means for increasing access to high-quality mental health care for harder-to-reach communities [53,54]. Virtual interventions have been found to reduce mental health stigma [52,53] and increase access to mental health treatment in historically underserved populations [52-54].

Therefore, we carried out a 2-site, pilot and feasibility randomized controlled trial of group IPT versus group CBT, delivered virtually, for adolescent girls with elevated anxiety and above-average BMI (institutional review board [IRB] number USUHS.2020‐048). In line with pilot study recommendations [55], main outcomes were the feasibility of recruitment, feasibility of research procedures, feasibility of retention, intervention fidelity, and acceptability of interventions.

Methods

Procedures

Enrollment for this pilot and feasibility randomized controlled trial was carried out at two sites: (1) the Uniformed Services University (USU) in Bethesda, Maryland, and (2) Colorado State University (CSU) in Fort Collins, Colorado. Participants were girls aged 12-17 years with elevated anxiety symptoms, as indicated by a total score of ≥32 on the State-Trait Anxiety Inventory for Children (STAI-C; [56]) and above-average BMI (≥75th percentile for age or sex). Exclusion criteria included any major medical conditions, pregnancy or breastfeeding, regular medication use likely to impact mood or weight (eg, insulin sensitizers, glucagon-like peptide-1 receptor agonists, antidepressants, and stimulants), current involvement in psychotherapy, and/or a psychiatric disorder or symptoms (eg, active suicidal ideation, behavior, or self-harm) that, in the opinion of the investigators, would impede competence, compliance, or otherwise hinder completion of the study. Exclusion criteria were chosen to ensure that adolescents who needed a higher level of care—beyond what could be provided in the context of a research study—were not included and any medications or interventions that could impact anxiety, eating behaviors, and/or cardiometabolic health were removed to mitigate confounding variables. If participants developed exclusionary criteria after randomization, they were withdrawn

XSL•FO
RenderX

if participation was counter-indicated for safety or validity in the opinion of the investigators.

Recruitment took place from February 2021 to August 2021. Letters and flyers were sent to parents of girls aged 12-17 years who resided within 60 miles of USU or CSU; letters were also mailed to local pediatricians and family physicians. The study was advertised on school and parent email listserves, in local newspapers, and on social media. The target sample size was 40 participants (20 per site), in line with recommendations for pilot studies in which the primary effect size in future efficacy trials is expected to be small to moderate [57].

Adolescents and their parents initially completed a phone prescreening to estimate eligibility. Adolescents who were interested and appeared eligible were scheduled for an in-person screening and baseline data collection visit at either USU or CSU. The screening visit included written consent and assent, body measurements and fasting blood draw, brief psychiatric and medical health history with the participant and parent or guardian, clinical assessment interview, and surveys. Medical health history included age of menarche and menstrual cycle status. Parents completed questionnaires about parental education (highest education attained) and subjective socioeconomic status using the MacArthur Scale of Subjective Social Status [58].

Questionnaires, such as the family history form, not previously validated with other populations, were developed by principal investigators and statisticians and piloted with study staff. Surveys were password-protected; passwords were known only by study staff and were completed via a secure electronic data-capture system, REDCap. Before completing any questionnaires, both parents and adolescents were informed of the length of the time of the surveys, where data would be stored (ie, REDCap), the names of the principal investigators, and the purpose of the questionnaires. As all surveys consisted of prespecified measures, randomization of questions and adaptive questionnaires were not used, and incomplete questionnaires were included in analysis.

Eligible participants were randomly assigned to IPT or CBT. Participants were randomized using a computerized number generator created by a statistician; only the statistician knew the sequence. Randomization was performed by the statistician only after eligibility was determined. Every consecutive block of 2 - 7 girls at each location was randomized to IPT or CBT. Randomization continued until a full cohort (4 - 7 adolescents per group arm) was created. Neither interventionists, by design, nor outcome assessors were blind to assignment; research staff performing phlebotomy were blinded.

All virtual groups were cofacilitated by a psychologist and a graduate student with clinical training; groups were facilitated via the Health Insurance Portability and Accountability Act–compliant, secure platform Zoom. Both facilitators conducted both the IPT and CBT groups. Reminder emails were sent 24 hours prior to each group. Sessions lasted 90 minutes and took place weekly after school hours for 12 weeks. Fidelity and adherence were monitored through weekly supervision meetings that included feedback on audio-recorded sessions from a clinical psychologist with IPT and CBT expertise. Both programs included one 90-minute and two 30-minute individual sessions at the beginning, middle, and end of the 12-week group program, respectively.

At the first individual session, a parent or guardian briefly joined (10 - 15 minutes) to review virtual group participation guidelines and problem-solve potential engagement barriers (eg, access to a device with camera and microphone in a private space). All participants received a text message reminding them to attend group 24 hours prior. Adolescents in both arms were encouraged to apply skills learned in group sessions to anxiety-provoking situations between sessions via home practice exercises facilitated through mailed binders containing handouts. Before each session, adolescents completed a mood-monitoring form, which assessed suicidality and self-harm; if either was endorsed, adolescents met with a psychologist for assessment and planning for safety monitoring.

Study flow, including time to enroll the total target sample, was tracked following CONSORT (Consolidated Standards of Reporting Trials) guidelines for pilot studies [59]. Recruitment feasibility was operationalized as enrollment of the target sample size (n=40) within 12 months. The percentage of eligible youth who enrolled was also calculated. Retention was tracked at each follow-up interval. Benchmarks for retention feasibility were defined as ≥80% retention at posttreatment or 12-week follow-up and ≥70% retention at 1-year follow-up. Feasibility of the study protocol was assessed by research staff completion of study flow checklists for standardized operating procedures at each interval. A threshold of ≥90% adherence to standardized procedures was expected, with ≤5% missing data for surveys, interviews, body measurements, blood pressure, and blood collection among completers at each interval.

Primary Aims

Primary study aims were recruitment and retention feasibility, study protocol or data collection feasibility, intervention fidelity, and program acceptability. Recruitment and retention feasibility were assessed through timely recruitment, attendance, and retention at 12-week and 1-year follow-up. Intervention fidelity was assessed via fidelity rating scales. Fidelity rating scales were created by 2 PhD-level mental or behavioral health clinician–researchers not involved in intervention delivery and/or supervision. The mental or behavioral health clinician–researchers met consistently with experts in IPT and CBT over 2 months to develop the fidelity measures. Fidelity rating scales were adapted from previously developed IPT and CBT program adherence measures provided by the study's primary investigators and were developed to reflect session-focused tasks specified in IPT and CBT intervention manuals. Fidelity measures were pilot tested using randomly chosen audio-recorded IPT and CBT sessions. Interrater reliability during pilot testing ranged between 0.95 and 1 (intraclass correlation coefficient 0.98; 95% CI 0.95 - 1.00).

Intervention fidelity was assessed through review of 20% of randomly selected audio-recorded IPT and CBT sessions by the 2 PhD-level mental or behavioral health clinician–researchers. In line with existing fidelity research [60], a benchmark of ≥80% adherence was set. The IPT fidelity checklist included items such as "Have ≥1 group member discuss an interpersonal

XSL•FO
RenderX

situation from the previous week and have the group ask questions; facilitate communication analysis if indicated." A CBT checklist example is: "Present a scenario where teens cope successfully with a problem and provide self-ratings." Items were rated 1=Yes, 0.5=Partially, or 0=No, with percentage fidelity calculated as the average adherence across all items and sessions per arm. Interrater reliability for 6 corated sessions was excellent (intraclass correlation coefficient 0.96; 95% CI 0.92-0.98).

To assess intervention acceptability, at 12-week follow-up, adolescents completed an adapted Treatment Acceptability Questionnaire [61], which measures likeability and credibility of IPT and CBT on a Likert scale (0=Not at all to 4=Definitely). Acceptability was defined as an average score ≥3 on likeability or credibility ratings within each arm. Acceptability was also characterized by attendance; ≥80% attendance at ≥10 of 12 group sessions was anticipated in both arms.

Secondary Aims

Secondary aims were to describe within-arm changes in anxiety symptoms, disinhibited eating (emotional eating and loss-of-control eating), BMI indices and fat mass percentage, and cardiometabolic health indicators including blood pressure, cholesterol, triglycerides, glucose, and hemoglobin A_{1c} (HbA_{1c}). Anxiety and eating behaviors were collected through online survey self-assessments, apart from the Eating Disorder Examination–Overeating Section [62], which was collected via interview by trained research staff. The STAI-C [56] assessed eligibility (total score ≥32; range 20‐60) and anxiety symptom severity. The STAI-C [56] has been validated for screening anxiety in youth with chronic health conditions with good psychometric properties [63]. Anxiety was measured at baseline and 1-year follow-up. Given the relevance of social anxiety in adolescents with elevated BMI, participants also completed the Social Phobia and Anxiety Inventory for Children [64], a 26-item self-report assessing somatic, cognitive, and behavioral aspects of social anxiety. Higher total scores (range 0‐52) indicate more symptoms. The Social Phobia and Anxiety Inventory for Children [64] is validated in children aged 8‐17 years, with good psychometric properties [65] and was administered at baseline, 12-week, and 1-year follow-ups.

The Emotional Eating Scale Adapted for Children and Adolescents [66] is a 25-item self-report assessing the urge to eat in response to negative affect. Items are measured on a 5-point scale [66], with higher scores reflecting a greater reported desire to eat when experiencing negative mood states [66]. Scores (range 0‐100) are summed. The Emotional Eating Scale Adapted for Children and Adolescents demonstrates good internal consistency and validity [67]. Loss-of-control eating was assessed via the Eating Disorder Examination–Overeating Section [62] administered by trained staff. Frequency of loss-of-control episodes in the past month was recorded as total episodes characterized by subjective inability to control eating any amount of food (subjective and/or objective binge-eating episodes) [62].

Height was measured in triplicate with a stadiometer (Model HM 200P; Charder Electronic COLTD) [68] and averaged. Height and weight were used to calculate BMI indices [69]. Fasting weight was measured on a calibrated digital scale connected to a Bod Pod [70]. Body fat percentage was measured via air displacement plethysmography using Bod Pod in a fasted state with minimal clothing and a swim cap [70].

Fasting blood samples were collected to measure glucose, cholesterol, triglycerides, and HbA_{1c}. Resting blood pressure was measured in triplicate with an automatic monitor (Model UA-789AC; LifeSource) and averaged. Blood samples were processed at the University of Colorado Anschutz Medical Campus Core Lab using standard procedures.

CONSORT study flow information and descriptive summary statistics were used to assess recruitment and retention feasibility, study protocol or data collection feasibility, intervention fidelity, and program acceptability. Group medians, interquartile ranges, and 95% CIs were reported at each measurement interval, including baseline, 12-week follow-up, and 1-year follow-up, to illustrate patterns within arms over time. These descriptive data include all participants who completed each specific time point, resulting in the possibility of different sample sizes at each time point. Multiple imputation was not used, given that it relies on patterns in the data and therefore may be noisy or unreliable given the small sample size per arm [71]. To describe changes in clinical health-related outcomes, change scores from baseline to 12-week follow-up and baseline to 1-year follow-up were examined in each arm (IPT and CBT) separately, and this analysis included participants who completed both time points needed for calculation of the change score (ie, baseline and 12-week; baseline and 1-year).

Given the pilot nature of the study, in which we were focused on the estimation of a signal, a paired-samples 2-tailed *t* test was used to compare with a nonzero benchmark. Paired-samples *t* tests were used for normally distributed outcomes, and Wilcoxon signed rank tests were used when outcomes deviated from normality. Cohen *d* was calculated to describe effect sizes for normally distributed outcomes, while *r* was used for nonnormal outcomes. To control for multiple comparisons, adjusted *P* values using the Benjamini-Hochberg false discovery rate are reported. Because behavior, anxiety, BMI indices, and cardiometabolic health indicators were each prespecified as dependent variables, we applied a family-wise false discovery rate, grouping 6 outcomes related to anxiety, 8 outcomes related to disinhibited eating, 8 outcomes related to BMI indices, and 36 outcomes related to cardiometabolic indicators. For acceptability ratings, Mann-Whitney *U* tests were used to test for potential group differences across domains of acceptability.

Ethical Considerations

The use of human subjects in this research was reviewed and approved by the USU IRB (number USUHS.2020‐048). Adolescents and parents or guardians were provided with informed consent and assent documents by trained study staff at the outset of the first screening visit. Informed consent documents included information on the nature and purpose of the study, study procedures including at-home device wear and group interventions, confidentiality, study benefits including compensation, study risks, alternative options (eg, therapy), and contact information if concerns arose. Adolescents were compensated for participating in various aspects of the study;

XSL•FO
RenderX

in total, they could earn up to US $917 if they completed all visits, at-home data collection, surveys, group sessions, and follow-up visits. All data collected were deidentified.

Study staff who administered surveys were trained by a psychologist in safety and risk assessment. Any participants who expressed active self-harm or suicidal ideation were referred to crisis numbers and provided with counseling referrals. Adverse events were monitored by the study principal investigators (blinded for review) and the site principal investigators in conjunction with oversight from an external safety officer and study physician (blinded for review). Adverse events were promptly reported to the principal investigators, and any serious adverse events were immediately reported to the USU IRB. The investigators prepared a quarterly report for the Data Safety Officer with enrollment, retention, adverse events, and deviations. Following recommended guidelines, serious and/or unanticipated adverse events were reported within 5 business days to the IRB and the Data Safety Officer. The safety officer reviewed the quarterly data and safety-monitoring reports and, if needed, recommended corrective action or reporting of out-of-range laboratory data to the participant and their parents or guardians.

Results

Participants

A total of 40 adolescents were enrolled and randomized within 7 months. Mean age, race or ethnicity, and weight status were comparable across IPT and CBT arms. Table 1 displays demographic characteristics, BMI status, and socioeconomic information by arm. Adolescents in IPT had greater baseline percentage adiposity than adolescents in CBT (38.3% vs 32.0%), with no other significant group differences at baseline in any other characteristic. There was a trend toward a group difference in the racial composition of each arm, with IPT having more White (70.0% vs 55.0%) and Black or African American (30.0% vs 10.0%) participants, and the CBT group having more Asian (5.0% vs 0.0%) participants, more participants reporting more than 1 race (20.0% vs 0.0%), and more participants who did not report their race (10.0% vs 0.0%). Most IPT participants (85.0%) and all CBT participants had started their menstrual cycle at baseline. Average age at menarche for participants enrolled in IPT was 11.3 (SD 1.3), which was similar to participants in CBT (mean 11.5, SD 0.83). Most participants in IPT and CBT, respectively, reported experiencing menstruation every 28 days (64.7% vs 75.0%).

XSL•FO
RenderX

Table . Participant baseline demographic, weight, and socioeconomic characteristics by arm. Parent education is the highest level of education of the parent responding to the questionnaire.

	IPT[a] (n=20)	CBT[b] (n=20)	*P* value
Demographic			
Age (years), mean (SD)	14.9 (1.8)	14.8 (1.8)	.90
Started menstrual cycle, % (n/N)	85 (17/20)	100 (20/20)	.23
Age at menarche (years), mean (SD)	11.3 (1.3)	11.5 (0.83)	.66
28-day cycle, % (n/N)	64.7 (11/17)	75 (15/20)	.72
Race, n (%)			
White	14 (70)	11 (55)	.05
Black or African American	6 (30)	2 (10)	
Asian	0 (0.0)	1 (5)	
More than 1 race	0 (0.0)	4 (20)	
Missing or chose not to report	0 (0.0)	2 (10)	
Ethnicity, n (%)			
Hispanic/Latinx	2 (10)	3 (15)	.63
Non-Hispanic or non-Latinx	18 (90)	17 (85)	
Weight and adiposity, n (%)			
Obesity, BMI ≥95th percentile	13 (65)	12 (60)	.52
Overweight, BMI 85th to 94th percentile	6 (30)	8 (40)	
Nonoverweight, 75th to 84th percentile	1 (5)	0 (0.0)	
Percent fat mass, mean (SD)	38.3 (10.5)	32.0 (6.3)	.03
Socioeconomic			
Parent education, n (%)			
Less than high school	0 (0.0)	0 (0.0)	.25
High school or GED[c]	0 (0.0)	1 (5)	
Partial college	4 (20)	1 (5)	
Associate’s degree	0 (0.0)	0 (0.0)	
Four-year college degree	13 (65)	10 (50)	
Advanced degree	3 (15)	6 (30)	
Missing or chose not to report	0 (0)	2 (10)	
Parent subjective social status, mean (SD)			
Relative to the United States	5.3 (1.1)	5.9 (1.7)	.16
Relative to one’s own community	5.3 (1.3)	5.7 (1.7)	.42

[a]IPT: interpersonal psychotherapy.

[b]CBT: cognitive behavioral therapy.

[c]GED: General Educational Development.

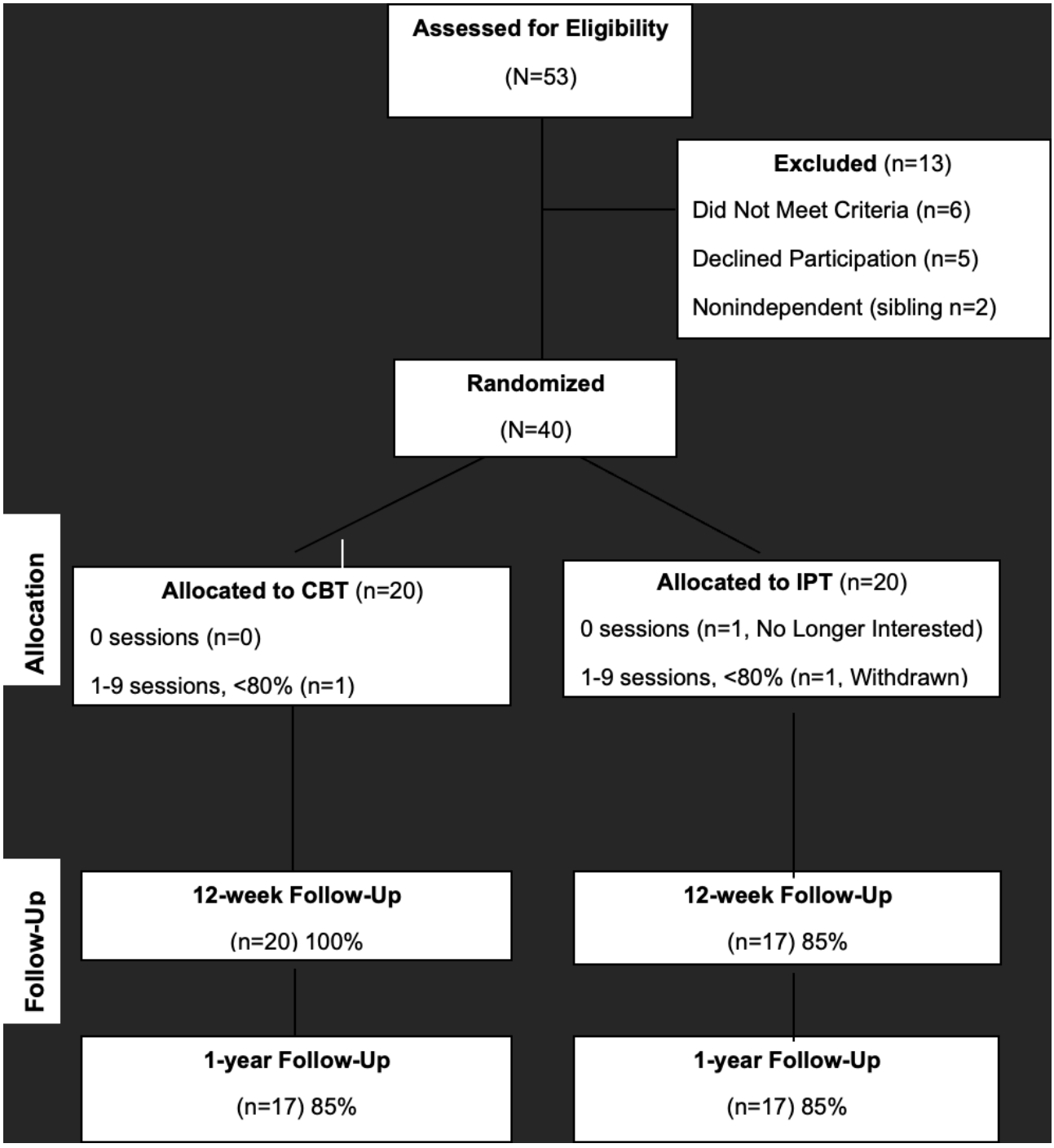
Primary Aims

Recruitment occurred from February 2021 to August 2021, with the total target enrollment of 40 reached in 7 months, corresponding to a rate of 5 - 6 enrolled participants per month across 2 sites. A total of 53 adolescents were assessed; 6 were ineligible. Of the 47 adolescents eligible after baseline screening, 40 (85.1%) chose to enroll. CONSORT study flow is depicted in Figure 1 (retention feasibility). Eighty-five percent (17/20) of the participants randomized to IPT completed the 12-week

XSL•FO
RenderX

follow-up. In IPT, 1 participant withdrew before starting due to loss of interest, 1 IPT participant was withdrawn due to an unexpected, serious adverse event unrelated to the protocol, and 1 participant was unable to complete the 12-week follow-up due to scheduling conflicts. Overall, 100% (20/20) of those randomized to CBT completed the 12-week follow-up. At the 1-year follow-up, 85% (17/20) of the IPT participants and 85% (17/20) of the CBT participants completed follow-up.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) study flow diagram for a web-based 2-site, 2-arm pilot and feasibility randomized controlled trial to improve cardiometabolic health in adolescent girls with elevated anxiety. CBT: cognitive behavioral therapy; IPT: interpersonal psychotherapy.



Based on staff completion of standardized procedure checklists at each time point, compliance was 98.2% (7.86/8 procedures) at screening or baseline, 98.1% (7.85/8 procedures) at 12-week follow-up, and 96.6% (7.73/8 procedures) at 1-year follow-up. Table 2 presents data missingness by measurement type (questionnaires or interviews, body measurements, and phlebotomy) for each interval by arm. Expert ratings of intervention adherence indicated acceptable fidelity to IPT (mean 78%, SD 11%) and high fidelity to CBT (mean 94%, SD 9%).

XSL•FO
RenderX

Table . Missingness of data for adolescent participants who completed each interval within arm over time[a].

	IPT[b], % (n/N)			CBT[c], % (n/N)		
	Baseline (n=20)	12 weeks (n=17)	1 year (n=17)	Baseline (n=20)	12 weeks (n=20)	1 year (n=17)
Questionnaire						
Treatment acceptability	N/A[d]	5.9 (1/17)	N/A	N/A	10.0 (2/20)	N/A
Anxiety	0.0 (0/20)	N/A	0.0 (0/17)	0.0 (0/20)	N/A	0.0 (0/17)
Social anxiety	5 (1/20)	5.9 (1/17)	0.0 (0/17)	5 (1/20)	15 (3/20)	0.0 (0/17)
Emotional eating	0.0 (0/20)	5.9 (1/17)	0.0 (0/17)	0.0 (0/20)	10 (2/20)	0.0 (0/17)
Interview						
Loss-of-control eating	0.0 (0/20)	0.0 (0/17)	5.9 (1/17)	0.0 (0/20)	5 (1/20)	5.9 (1/17)
Body measures						
BMI indices	0.0 (0/20)	0.0 (0/17)	0.0 (0/17)	0.0 (0/20)	0.0 (0/20)	0.0 (0/17)
Body fat[e]	0.0 (0/20)	0.0 (0/17)	0.0 (0/17)	5 (1/20)	5 (1/20)	0.0 (0/17)
SBP/DBP[f]	5 (1/20)	5.9 (1/17)	0.0 (0/17)	0.0 (0/20)	0.0 (0/20)	0.0 (0/17)
Phlebotomy						
Lipids[g]	20 (4/20)	17.6 (3/17)	23.5 (4/17)	15 (3/20)	20 (4/20)	23.5 (4/17)
HbA_{1c}[h]	25 (5/20)	17.6 (3/17)	11.8 (2/17)	20 (4/20)	30 (6/20)	17.6 (3/17)
Glucose	20 (4/20)	17.6 (3/17)	23.5 (4/17)	15 (3/20)	20 (4/20)	23.5 (4/17)

[a]Values are % missing (missing data/n completed interval).

[b]IPT: interpersonal psychotherapy.

[c]CBT: cognitive behavioral therapy.

[d]N/A: not applicable because not collected at this interval.

[e]Body fat: air displacement plethysmography/Bod Pod.

[f]SBP/DBP: systolic and diastolic blood pressure.

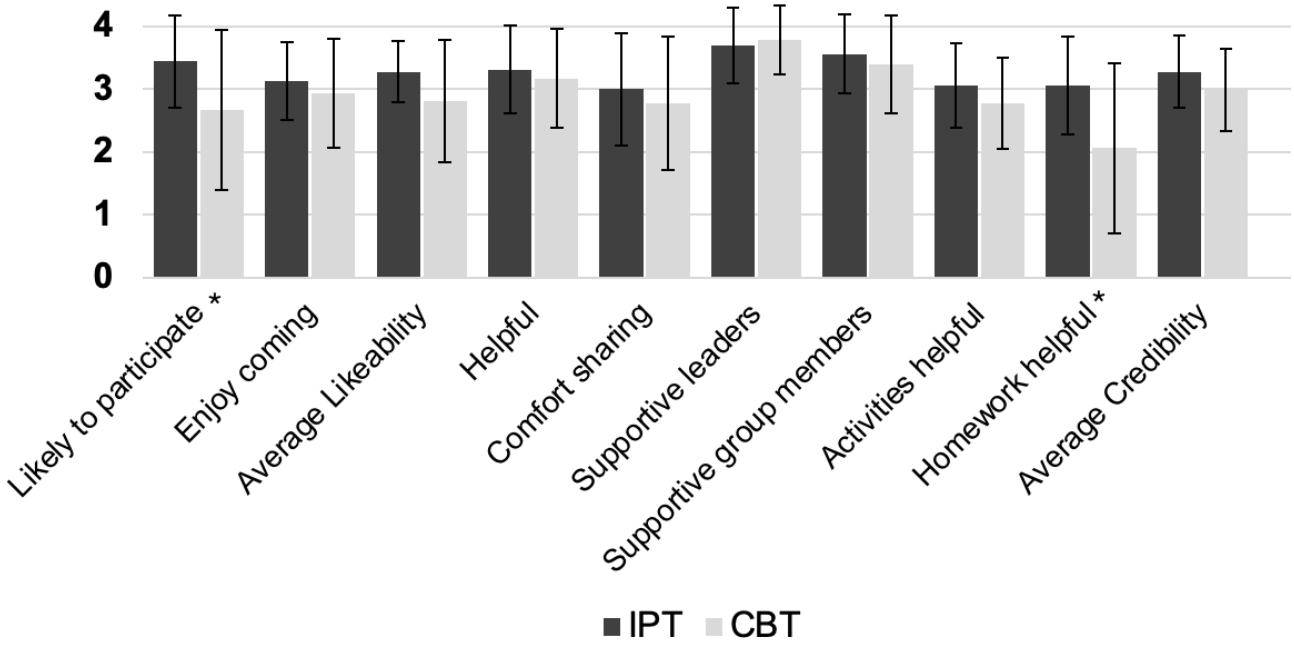
[g]Lipids: triglycerides and cholesterol.

[h]HbA_{1c}: hemoglobin A_{1c}.

Figure 2 shows average acceptability ratings by arm. Acceptability (likability and credibility) survey results exclude participants with missing survey data, resulting in 16 adolescents completing surveys in IPT and 18 in CBT. Nonparametric tests of group differences revealed higher ratings of the helpfulness of at-home activities in the IPT group (U=204.00, mean rank=21.25 in IPT vs 14.17 in CBT; P=.04). There was also a trend where IPT participants were more likely to report that they would participate in the program if they could do it over (U=193.50, mean rank 20.59 vs 14.75 in CBT; P=.09). All Mann-Whitney comparisons should be interpreted with caution, however, given the small sample size.

XSL•FO
RenderX

Figure 2. Average acceptability and credibility ratings by intervention arm for a web-based 2-site, 2-arm pilot and feasibility randomized controlled trial to improve cardiometabolic health in adolescent girls with elevated anxiety. Acceptability ratings among adolescent girls with elevated anxiety and above-average weight randomized to 12-week, virtual group IPT or 12-week, virtual group CBT. Values are mean (SD) on 0-4 scale, with 0=Lowest likability or credibility and 4=Highest likability or credibility. *Statistically significant and reported in the aforementioned text. The values for likely to participate are as follows: CBT: mean 2.67 (SD 1.28); IPT: mean 3.44 (SD 0.18); and *t* test: t_{32}=−2.12; *P*=.04. The values for homework helpful are as follows: CBT: mean 2.17 (SD 1.34); IPT: mean 3.13 (SD 0.81); and *t* test: t_{32}=−2.45; *P*=.01. CBT: cognitive behavioral therapy; IPT: interpersonal psychotherapy.



Descriptively, in IPT, 87.5% (14/16) of adolescents reported that they would “mostly” or “definitely” complete the program again (3 or 4 on a 0 - 4 scale), 87.5% (14/16) found the group “very” to “definitely” enjoyable, and 87.5% (14/16) thought the group would be “mostly” or “extremely” helpful for peers. Most adolescents in IPT felt “very” or “extremely” comfortable sharing and asking questions (12/16, 75%), viewed facilitators (15/16, 93.8%) and group members (15/16, 93.8%) as supportive, found group activities “very” or “extremely” helpful (13/16, 81.3%), and saw home practice exercises as “very” or “extremely” useful (12/16, 75%).

In CBT, most adolescents (11/18, 61.1%) reported that they would “mostly” or “definitely” complete the program again, 66.7% (12/18) found the group “very” to “definitely” enjoyable, and 72.2% (13/18) thought the group would be “mostly” or “extremely” helpful for their peers. Many adolescents also felt comfortable sharing and asking questions (13/18, 72.2%). Similar to IPT, adolescents viewed facilitators (17/18, 94.4%) and group members (17/18, 94.4%) as supportive. Most adolescents in CBT found group activities “very” or “extremely” helpful (13/18, 72.2%), but a minority saw home practice exercises as “very” or “extremely” useful (8/18, 44.4%).

In the intent-to-treat sample, 90% (18/20) in IPT and 95% (19/20) in CBT received ≥80% of the intervention dose (≥10 of 12 group sessions). Median attendance was 11 sessions in IPT and 12 in CBT. One IPT participant withdrew before starting due to loss of interest. One IPT participant was withdrawn due to an unexpected, serious adverse event unrelated to the protocol; the participant reported worsening mental health symptoms that necessitated hospitalization and precluded ongoing research participation. Excluding withdrawn participants, mean attendance was 11.00 (SD 1.46) sessions in IPT and 11.55 (SD 1.36) sessions in CBT.

Secondary Aims

Key clinical outcomes at each interval by arm are detailed in Table 3. Change scores in anxiety, disinhibited eating, BMI indices, and adiposity, as well as cardiometabolic health indices adjusted for multiple comparisons and within-arm effect sizes, are shown in Table 4. From baseline to 12-week follow-up, adolescents enrolled in CBT (but not IPT) showed improvement in emotional eating and adolescents enrolled in IPT (but not CBT) demonstrated improvement in BMI percentile. From baseline to 1-year follow-up, adolescents enrolled in IPT (but not CBT) showed improvement in general anxiety, social anxiety, and HbA_{1c}, while adolescents in CBT (but not IPT) demonstrated improvements in emotional eating. Both arms demonstrated significant improvement in BMI percentile at 1-year follow-up. No significant changes were observed in loss-of-control eating, BMI, fat mass percentage, blood pressure, triglycerides, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, or glucose at 1-year follow-up in either arm.

XSL•FO
RenderX

Table . Anxiety, emotional eating, BMI or adiposity, and cardiometabolic health indices by intervention arm across baseline, 12-week, and 1-year intervals. Denominators reflect participants who completed assessments at each interval.

	IPT[a]						CBT[b]					
	Baseline		12 weeks		1 year		Baseline		12 weeks		1 year	
Measure	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)
Anxiety and disinhibited eating												
Anxiety	20	41.5 (35.3-46.0)	—[c]	NA[d]	17	*35.0* (*29.0-42.0*[e])	20	39.5 (35.5-43.8)	—	NA	17	38.0 (30.5-42.0)
Social anxiety	19	24.0 (16.0-35.0)	16	17.50 (14.0-25.0)	17	*17.0* (*9.5-24.5*)	19	23.0 (15.0-32.0)	17	23.0 (12.0-28.5)	17	11.0 (4.0-25.0)
Emotional eating	20	1.5 (1.1-1.9)	16	1.3 (0.7-1.7)	17	1.2 (0.5-1.6)	20	1.3 (1.0-2.1)	18	*0.3* (*0.1-0.8*)	17	*0.5* (*0.2-1.1*)
Loss of control	20	0.0 (0-3)	17	0.0 (0-1)	16	0.0 (0-0.75)	20	0.0 (0-0)	19	0.0 (0-0)	16	0.0 (0-0)
BMI and adiposity												
BMI, kg/m^2	20	31.4 (25.6-37.7)	17	32.0 (25.6-38.8)	17	32.6 (24.4-37.9)	20	28.2 (26.0-31.0)	20	28.1 (26.6-31.3)	17	30.0 (25.9-32.2)
BMI, percentile	20	97.0 (91.0-99.0)	17	*98.0*[f] (*89.5-99.0*)	17	*98.0* (*84.5-99.0*)	20	95.0 (93.0-97.8)	20	95.0 (94.0-97.0)	17	*95.0* (*90.0-98.0*)
Fat mass, percentage	20	39.4 (28.3-46.6)	17	40.6 (30.0-43.7)	17	41.0 (30.9-47.7)	19	30.8 (27.0-37.5)	19	30.0 (27.8-32.3)	17	32.1 (27.2-36.6)
Cardiometabolic health												
SBP[g], mm Hg	19	111.3 (108.7-122.7)	16	114.5 (106.8-125.2)	17	111.0 (103.3-126.3)	20	119.8 (111.3-126.0)	20	123.5 (110.8-126.0)	17	111.3 (108.2-124.3)
DBP[h], mm Hg	19	78.7 (69.3-84.7)	16	76.5 (72.3-83.7)	17	67.7 (64.7-77.6)	20	76.0 (70.1-83.8)	20	79.7 (75.0-83.8)	17	72.0 (63.2-84.0)
Triglycerides, mg/dL	16	71.5 (59.0-97.0)	14	67.5 (51.8-96.5)	13	52.0 (44.5-96.0)	17	103.0 (69.0-128.0)	16	97.0 (68.5-152.5)	13	84.0 (70.5-119.5)
Total cholesterol, mg/dL	16	155.5 (138.0-165.3)	14	151.5 (138.0-162.8)	13	143.0 (126.5-159.5)	17	151.0 (125.0-165.5)	16	152.5 (130.3-166.8)	13	144.0 (124.5-174.5)
HDL[i] cholesterol, mg/dL	16	46.0 (39.0-49.0)	14	45.5 (38.5-49.3)	13	41.0 (38.00-53.5)	17	41.0 (36.5-49.0)	16	45.0 (40.8-50.8)	13	42.0 (35.5-46.0)
LDL[j] cholesterol, mg/dL	16	104.0 (87.3-117.3)	14	98.0 (88.0-113.0)	13	94.0 (75.5-105.5)	17	109.0 (74.0-116.0)	16	94.0 (75.0-114.5)	13	94.0 (72.5-118.0)
HbA_{1c}[k]	15	5.6 (5.4-5.7)	14	5.6 (5.4-5.7)	15	*5.2* (*5.1-5.5*)	16	5.5 (5.3-5.8)	16	5.7 (5.4-5.9)	14	5.3 (5.0-5.5)
Glucose, mg/dL	16	84.5 (78.3-90.8)	14	83.5 (77.8-87.0)	13	83.0 (80.0-92.0)	17	83.0 (81.0-85.0)	16	85.0 (81.5-88.0)	13	85.0 (83.5-89.0)

XSL•FO
RenderX

[a]IPT: interpersonal psychotherapy.

[b]CBT: interpersonal psychotherapy.

[c]Not available because not collected at this interval.

[d]N/A: not applicable because not collected at this interval.

[e]Values in italics are statistically significant.

[f]Time point specific summary statistics in Table 3 are based upon complete data (n=20, 17 at baseline and 12 weeks, respectively), whereas estimated difference scores in Table 4 hinge on complete information across both time points (n=17).

[g]SBP: systolic blood pressure.

[h]DBP: diastolic blood pressure.

[i]HDL: high-density lipoprotein.

[j]LDL: low-density lipoprotein.

[k]HbA_{1c}: hemoglobin A_{1c}.

XSL•FO
RenderX

Table . Within-condition changes and effect sizes within interpersonal psychotherapy and cognitive behavioral therapy arms using complete case analysis[a]. As this is a pilot study that does not formally compare groups, the authors caution against comparing treatments.

Characteristics and arms	n	*t* test (*df*) /*Z*[b,c]	*P* value	FDR[d]-adjusted *P* value	Mean/median Δ[e]	95% CI Δ	*d*/*r*[f,g] effect size
Anxiety							
Δ Baseline to 1 year							
IPT[h]	17	−2.91 (16)	.01	.03	−5.18	−8.95 to −1.40	−0.71
CBT[i]	17	−2.29 (16)	.04	.08	−3.76	−7.24 to −0.28	−0.56
Social anxiety							
Δ Baseline to 12 weeks							
IPT	15	−1.94 (14)	.07	.08	−5.53	−11.66 to 0.60	−0.50
CBT	16	−1.47 (15)	.16	.16	−2.38	−5.83 to 1.08	−0.37
Δ Baseline to 1 year							
IPT	16	−2.90 (15)	.01	.03	−7.13	−12.37 to −1.88	−0.72
CBT	16	−2.14 (15)	.05	.08	−6.75	−13.47 to −0.03	−0.54
Emotional eating							
Δ Baseline to 12 weeks							
IPT	16	−2.25 (15)	.04	.08	−0.25	−0.48 to −0.01	−0.56
CBT	18	−4.45 (17)	<.001	.01	−0.82	−1.21 to −0.43	−1.05
Δ Baseline to 1 year							
IPT	17	−1.89 (16)	.08	.13	−0.25	−0.54 to 0.03	−0.46
CBT	17	−2.88 (16)	.01	.04	−0.68	−1.18 to −0.18	−0.70
Loss-of-control eating							
Δ Baseline to 12 weeks							
IPT	17	−1.47	.14	.19	0.00	N/A[j]	−0.52
CBT	19	0	≥.99	>.99	0.00	N/A	0.00
Δ Baseline to 1 year							
IPT	16	−2.09	.04	.08	0.00	N/A	−0.70
CBT	16	−0.82	.41	.46	0.00	N/A	−0.47
BMI (kg/m^2)							
Δ Baseline to 12 weeks							
IPT	17	−0.55 (16)	.59	.76	−0.18	−0.86 to 0.51	−0.13
CBT	20	0.36 (19)	.72	.76	0.09	−0.40 to 0.57	0.08
Δ Baseline to 1 year							
IPT	17	−0.31 (16)	.76	.76	−0.16	−1.28 to 0.95	−0.08
CBT	17	0.69 (16)	.50	.76	0.25	−0.53 to 1.03	0.17
BMI percentile							
Δ Baseline to 12 weeks							
IPT	17	−2.53	.01	.04	−1.00[e]	N/A	−0.80
CBT	20	−1.24	.22	.44	0.00	N/A	−0.39
Δ Baseline to 1 year							
IPT	17	−2.85	.004	.03	−1.00	N/A	−0.82
CBT	17	−2.29	.02	.05	−1.00	N/A	−0.63

XSL•FO
RenderX

Characteristics and arms	n	*t* test (*df*) /*Z*[b,c]	*P* value	FDR[d]-adjusted *P* value	Mean/median Δ[e]	95% CI Δ	*d*/*r*[f,g] effect size
Fat mass (%)							
Δ Baseline to 12 weeks							
IPT	17	−1.60 (16)	.13	.64	−2.09	−4.86 to .68	−0.39
CBT	19	−0.94 (18)	.36	.78	−1.52	−4.94 to 1.90	−0.21
Δ Baseline to 1 year							
IPT	17	1.38 (16)	.19	.64	1.50	−0.80 to 3.78	0.34
CBT	16	−0.20 (15)	.84	.98	−0.25	−2.93 to 2.42	−0.05
SBP[k] (mm Hg)							
Δ Baseline to 12 weeks							
IPT	16	−0.70 (15)	.49	.83	4.00	N/A	−0.18
CBT	20	−0.90 (19)	.37	.78	0.17	N/A	−0.20
Δ Baseline to 1 year							
IPT	16	−0.65 (15)	.53	.83	−2.23	−9.56 to 5.10	−0.16
CBT	17	−1.26 (16)	.23	.64	−3.37	−9.11 to 2.37	−0.30
DBP[l] (mm Hg)							
Δ Baseline to 12 weeks							
IPT	16	−0.93 (15)	.37	.78	−5.88	−19.28 to 7.52	−0.23
CBT	20	0.65 (19)	.52	.83	1.55	−3.44 to 6.54	0.15
Δ Baseline to 1 year							
IPT	16	−2.77 (15)	.01	.12	−11.00	−19.48 to −2.52	−0.69
CBT	17	−1.28 (16)	.22	.64	−3.67	−9.73 to 2.40	−0.31
Triglycerides (mg/dL)							
Δ Baseline to 12 weeks							
IPT	13	−0.13 (12)	.90	.98	−0.62	−10.60 to 9.36	−0.04
CBT	14	1.53 (13)	.15	.64	11.00	−4.52 to 26.52	0.41
Δ Baseline to 1 year							
IPT	12	−1.33 (11)	.21	.64	−8.75	−23.23 to 5.73	−0.38
CBT	11	−0.70 (10)	.50	.83	−6.09	−25.40 to 13.21	−0.21
Total cholesterol (mg/dL)							
Δ Baseline to 12 weeks							
IPT	13	−0.45 (12)	.66	.92	−1.23	−7.19 to 4.73	−0.13
CBT	14	−0.37 (13)	.72	.92	−0.93	−6.30 to 4.45	−0.10
Δ Baseline to 1 year							
IPT	12	−2.31 (11)	.04	.29	−10.50	−20.53 to −0.47	−0.67
CBT	11	−1.07 (10)	.31	.78	−5.64	−17.35 to 6.08	−0.32
HDL[m] (mg/dL)							
Δ Baseline to 12 weeks							
IPT	13	0.00 (12)	≥.99	>.99	0.00	−2.47 to 2.47	0.00
CBT	14	−0.33 (13)	.75	.92	−0.36	−2.73 to 2.01	−0.09
Δ Baseline to 1 year							
IPT	12	0.77 (11)	.46	.83	0.92	−1.69 to 3.52	0.22

XSL•FO
RenderX

Characteristics and arms	n	*t* test (*df*) /*Z*[b,c]	*P* value	FDR[d]-adjusted *P* value	Mean/median Δ[e]	95% CI Δ	*d*/*r*[f,g] effect size
CBT	11	0.00 (10)	≥.99	>.99	0.00	−3.69 to 3.69	0.00
LDL[n] (mg/dL)							
Δ Baseline to 12 weeks							
IPT	13	0.15 (12)	.88	.98	0.31	−4.16 to 4.77	0.04
CBT	14	−0.33 (13)	.75	.92	−0.86	−6.51 to 4.79	−0.09
Δ Baseline to 1 year							
IPT	12	−1.73 (11)	.11	.64	−9.50	−21.59 to 2.59	−0.50
CBT	11	−1.32 (10)	.22	.64	−5.27	−14.17 to 3.62	−0.40
HbA_{1c}[o]							
Δ Baseline to 12 weeks							
IPT	12	0.59 (11)	.57	.86	0.03	−0.09 to 0.16	0.17
CBT	13	2.29 (12)	.04	.29	0.12	0.01 to 0.23	0.63
Δ Baseline to 1 year							
IPT	12	−5.75 (11)	.001	.04	−0.35	−0.48 to −0.23	−1.66
CBT	11	−3.02 (10)	.01	.12	−0.23	−0.39 to −0.06	−0.91
Fasting glucose (mg/dL)							
Δ Baseline to 12 weeks							
IPT	13	0.30 (12)	.77	.92	0.54	−3.32 to 4.40	0.08
CBT	14	−0.39 (13)	.70	.92	−0.64	−4.18 to 2.89	−0.11
Δ Baseline to 1 year							
IPT	12	0.65 (11)	.53	.83	1.00	−2.37 to 4.37	0.19
CBT	11	0.09 (10)	.93	.98	0.18	−4.26 to 4.62	0.03

[a]Values reported in the mean/median Δ column are descriptive; Wilcoxon signed-rank tests evaluate median change.

[b]*t*: *t* test,

[c]*Z*: Wilcoxon signed rank test,

[d]FDR: false discovery rate.

[e]Time point specific summary statistics in Table 3 are based upon complete data (n=20, 17 at baseline and 12 weeks, respectively), whereas estimated difference score in Table 4 hinges on complete information across both time points (n=17).

[f]*d*: effect size for *t* tests.

[g]*r*: effect size for Wilcoxon signed rank tests.

[h]IPT: interpersonal psychotherapy..

[i]CBT: cognitive behavioral therapy.

[j]N/A: not applicable because measure not collected at this interval.

[k]SBP: systolic blood pressure.

[l]DBP: diastolic blood pressure.

[m]HDL: high-density lipoprotein.

[n]LDL: low-density lipoprotein.

[o]HbA_{1c}: hemoglobin A_{1c}.

Discussion

Principal Findings

This study was the first to directly pilot the feasibility, acceptability, and exploratory outcomes on anxiety and cardiometabolic health markers of web-based, group IPT and CBT for adolescent girls with above-average weight and elevated anxiety symptoms. The target sample size of 40 adolescents was achieved across 2 sites in 7 months, with an enrollment rate of 5 - 6 participants per month. Of eligible adolescents, 85% (40/47) elected to enroll. Total retention was 93% (37/40) at 12-week follow-up and 85% (34/40) at 1-year follow-up. Specifically, at 12-week follow-up, 95% (19/20) of participants from USU and 90% (18/20) of participants at CSU completed the posttreatment visit (χ^2_2=1.3; *P*=.55). At both

XSL•FO
RenderX

sites, 85% (17/20) of participants completed 1-year follow up (χ^2_1=0.00; *P*>.99). Protocol adherence by research staff was high (>96%) across time points. Data missingness ranged from 0% to 30%, with no missing height or weight data. Questionnaire missingness ranged from 0% to 15%, and interview and body fat measurements missingness never exceeded 5.9%. Cardiometabolic measures requiring phlebotomy had the greatest missingness (15% - 25% at baseline, 11.8%-30% at follow-up). Intervention fidelity ratings indicated high CBT adherence (>90%); IPT adherence was just below the 80% threshold. Overall acceptability was high, with >90% (n≥17) receiving at least 80% of intervention dosage in both arms. Home practice assignments were rated as the least helpful aspect of both IPT and CBT web-based group interventions.

Principal Findings in the Context of Prior Work

Recruitment success likely reflects multifaceted strategies (letters, flyers, social media, and listserves) and the offer of evidence-based mental health interventions in both arms rather than a nonactive comparator, as reflected in qualitative work on research recruitment with adolescents [72] and similar to other trials comparing 2 evidence-based interventions [73]. Furthermore, the inclusion criteria of elevated anxiety symptoms and BMI ≥75th percentile, rather than overweight (≥85th percentile) or obesity (≥95th percentile), were appealing as they emphasized anxiety and stress prevention rather than weight loss, consistent with prior qualitative work [74]. Previous research indicates that effective retention strategies include flexible scheduling, easy interactions with research staff, and frequent contact [75]. These strategies were reflected in this trial, as staff offered flexible scheduling and focused on ongoing engagement via calls, birthday cards, and newsletters.

Few studies have evaluated therapist adherence to IPT and CBT simultaneously. Ekeblad and colleagues [44] compared IPT and CBT fidelity using the Collaborative Study Psychotherapy Rating Scale—6 [76]. Adherence was found to be at or above 80% [45]. The Collaborative Study Psychotherapy Rating Scale—6 was designed to assess IPT with *adult individuals* rather than *groups of adolescents,* preventing its use in this study. However, the adherence found in both studies is similar to the fidelity findings in the current protocol. We speculate that CBT's greater structure likely supported higher adherence, whereas IPT's flexibility in interpersonal content demands facilitator spontaneity and responsiveness to adolescent affect. The most missed IPT component included session summaries in favor of increased time in role-plays, highlighting training needs to balance fidelity and therapist competence. Competence in the context of fidelity refers to the idea that therapists can draw on a wide range of interpersonal and emotional skills to address challenges and respond to participants [77]. Below desired (ie, 78% rather than 80%) IPT fidelity ratings may reflect a greater emphasis on evaluation using a structured protocol rather than therapist competence and responsiveness to adolescent needs in IPT. Notably, adolescents reported greater enjoyment of IPT, potentially reflecting its personalized nature. Both IPT and CBT were led by the same facilitators, suggesting that this effect is not attributable to facilitator differences. The current results support the importance of measuring competence and personalizing treatment [78].

Despite adolescents' report that homework was the least helpful aspect of both groups, homework predicts clinical improvement [79]. Lower perceived helpfulness of homework may relate to negative attitudes toward academic homework and the burden of therapeutic tasks [80]. Additionally, adolescents diagnosed with anxiety may struggle to complete homework due to forgetfulness or misunderstandings of the most effective way to implement skills [81].

Overall, data missingness appears similar to other clinical trials involving phlebotomy with adolescents [82,83]. Despite this, virtual delivery of both group programs was successful. Virtual group formats are convenient for busy adolescents, minimizing travel burden and fostering peer connection [84]. Adolescents rated groups as enjoyable, comfortable, and helpful, with facilitators and peers perceived as supportive, similar to other group-based formats for adolescents [85].

Secondary Findings

Moderate to large reductions in general and social anxiety were observed at 1-year follow-up for adolescents enrolled in IPT, with no significant social anxiety changes at 12-week or postintervention follow-up. This pattern may indicate that treatment effects develop over time. Immediately post intervention, adolescents enrolled in IPT demonstrated significant reductions in BMI percentile, as well as at 1-year follow-up. Additionally, adolescents enrolled in CBT experienced significant change in both emotional eating and BMI percentile at 1 year postintervention follow-up. These preliminary and exploratory findings provide an initial signal that IPT and CBT may influence mechanisms that trigger disinhibited eating, yet this hypothesis would need to be explicitly tested in larger samples powered to study mechanistic effects. The positive pattern of change in eating behavior in CBT, although highly preliminary, is notable in the context of prior data that disinhibited eating often worsens, and behavioral weight interventions typically show rebound effects posttreatment [86]. Some metabolic outcomes improved descriptively, with CBT showing improvements in BMI percentile and IPT demonstrating HbA_{1c} improvements at 1 year. While highly preliminary, these initial data suggest that web-based group anxiety interventions have the potential to support healthier cardiometabolic trajectories in at-risk adolescent girls.

Secondary Findings in Context of Prior Work

Consistent with previous literature [87], both IPT and CBT resulted in moderate to large reductions in anxiety symptoms. Extant research also supports the use of mental health interventions to address the potential mediators of physical health conditions [37]. For example, mental health interventions have been found to improve both anxiety symptoms and cardiometabolic health in adolescents with excess weight [86].

XSL•FO
RenderX

Strengths and Limitations

Strengths of this study include methodological rigor with strong recruitment, retention, protocol adherence, intervention attendance, and acceptability. Limitations include the following.

First, this study has limited power. Notably, the study was not powered to assess comparative effects; within-arm changes should be interpreted cautiously, given the small sample size. Likewise, as this study was a pilot of 2 active interventions without an inactive control arm, any within-arm improvements cannot be attributed to the specific modalities tested; it is possible that changes were influenced by natural maturation, nonspecific therapeutic attention effects, and/or regression to the mean.

Second, missing data were not uniform across outcomes and were concentrated in phlebotomy-derived cardiometabolic measures. Whereas questionnaire, interview, and anthropometric outcomes were largely complete, blood-based measures had 15% - 30% missingness across arms and time points, exceeding the study's a priori feasibility benchmark. Accordingly, exploratory findings for metabolic (eg, HbA_{1c}) outcomes should be interpreted cautiously, as missingness likely reduced precision, effective sample size, and power to detect small to-moderate effects. Because this pilot was not designed or powered to formally evaluate missing data mechanisms, the extent of potential bias remains uncertain. Exploratory comparisons of participants with versus without follow-up phlebotomy data on baseline characteristics showed no meaningful baseline differences, suggesting limited evidence of systematic missingness based on observed variables. However, the group of participants with missing baseline phlebotomy data had higher baseline BMI percentile (mean 97.70, SD 0.64 vs mean 94.12, SD 4.90; P=.01) and a greater proportion of Hispanic and Latinx participants (40% vs 6%, Fisher exact test P=.03) versus the group with baseline phlebotomy data. Therefore, we cannot rule out the possibility of systematic missingness at baseline based on higher BMI percentile and ethnicity. In future studies, it will be important to ensure that baseline measures of phlebotomy are collected across BMI strata and race and ethnicity, particularly in youth with higher BMI and Hispanic Latinx ethnicity, addressing any potential barriers in youth with more severe obesity or Hispanic or Latinx ethnicity.

Third, although IPT intervention fidelity did not reach prespecified thresholds, most of the feasibility and acceptability markers did, suggesting strong support for feasibility in 7 of 9 dimensions with room to improve in both phlebotomy and IPT fidelity. As IPT was just below the 80% benchmark and we did not measure competence, it is possible that the therapeutic dosage was suboptimal. Nevertheless, the most common missing fidelity requirement was summarizing the session, suggesting that core content was administered to fidelity.

Fourth, while social anxiety symptoms were assessed at baseline and all follow-up intervals, general anxiety symptoms were assessed only at baseline and 1 year as part of an effort to limit participant questionnaire fatigue and given the study's focus on feasibility and acceptability as primary outcomes. Additional limitations include that homework completion was not objectively measured.

Fifth, participants could not be blinded, which increased the risk for type I error due to the multiplicity of outcomes. Sixth, the absence of qualitative data limits participant perspective on intervention experience. Seventh, the sample was demographically homogenous, thereby limiting external validity and generalizability of the feasibility and acceptability findings to lower-income or minority populations. Eighth, reductions in loss-of-control eating episodes must be considered cautiously, as not all adolescents had loss of control at baseline and frequencies of loss-of-control episodes were low. Finally, there was a group difference in percentage adiposity at baseline, with the IPT arm having higher percentage adiposity than the CBT arm; it is possible that outcomes pertaining both to adiposity and other variables may relate to this baseline characteristic.

Future Work

Future fully powered trials should be designed to test comparative efficacy between IPT and CBT, using a prespecified primary end point and adequate sample size. In the present pilot, within-arm effect sizes for key outcomes were generally in the moderate to large range. For planning purposes, anchoring to a moderate effect (d=0.50; α=.05, 80% power), a future trial would require approximately 64 participants per arm with complete data, corresponding to approximately 93 - 99 participants per arm after accounting for approximately 20% - 30% attrition and biospecimen missingness. Larger samples would be required to adequately detect between-group differences in treatment response. Furthermore, future research should engage adolescent interest holders to identify barriers and facilitators to homework adherence and explore digital or technological supports as enhancement to web-based group interventions.

Therefore, areas of refinement include phlebotomy procedures, specifically, planning for larger sample sizes to accommodate biospecimen missingness and prospectively documenting reasons for missed or unsuccessful blood draws, IPT training or supervision related to adherence and competence, sufficient retention to confirm the impact of IPT versus CBT on cardiometabolic health for adolescent girls, and including nonactive control groups. Specifically, future trials should anticipate missing data from phlebotomy and consider specialized personnel or noninvasive alternatives (eg, ultrasound) to reduce missingness; use of finger pricks or nonfasting measures would also be expected to improve data completeness [83]. Fully powered efficacy trials should also aim to assess the mediational pathways of the mechanisms of both IPT and CBT, evaluate generalized anxiety at all follow-up intervals to better characterize trajectories of anxiety, plan for greater missingness in power estimations, and account for pubertal changes during the follow-up period. Additionally, future trials should test the impact of interventions for girls with more frequent loss-of-control eating and/or binge eating disorder.

XSL•FO
RenderX

Conclusions

This pilot trial provides initial support for feasibility and acceptability of virtual group IPT and CBT for adolescent girls with anxiety symptoms who also are at risk for future, preventable cardiometabolic disease due to above-average BMI. Specifically, 7 out of 9 a priori benchmarks demonstrated feasibility (ie, recruitment, retention, protocol adherence, nonphlebotomy data missingness, CBT intervention fidelity, and both IPT and CBT intervention likability and credibility). The results also pointed to 2 areas to refine (ie, phlebotomy missingness and IPT intervention fidelity) for a future trial.

Given the high prevalence of anxiety among adolescent girls with excess weight and evidence that mental health interventions have the potential to improve not only mental health but also behavioral and cardiometabolic outcomes, targeting anxiety represents a promising and novel prevention strategy. Routine pediatric clinical care, particularly care designed to address excess weight and its comorbidities (eg, lifestyle medicine), would likely benefit from integrated group mental health interventions to address mental health concerns such as anxiety. Web-based group interventions may offer a scalable and accessible approach to addressing these intersecting risks and will benefit from continued interdisciplinary collaboration across psychology, pediatrics, and cardiometabolic health. Yet, the current results must be considered preliminary. Building on the feasibility and acceptability demonstrated in this pilot, a fully powered randomized controlled trial is warranted to evaluate the efficacy and mechanisms of virtual group IPT and CBT in this population.

Acknowledgments

The authors gratefully acknowledge Retired Captain Jeffrey Quinlan, MD, FAAFP, for serving as the Data Safety Officer for this study, and the authors extend their sincere appreciation to the adolescent participants and their families who volunteered their time and made this research possible. The opinions and assertions expressed herein are those of the authors and do not reflect the official policy or position of the Uniformed Services University or the Department of Defense. Artificial intelligence (Microsoft Copilot) was used solely to format the CONSORT (Consolidated Standards of Reporting Trials) diagram from provided numbers.

Funding

Funding for this study was provided by Defense Health Agency grants HU000119200029 and HU00012420018 (Haigney). The study was funded for 5 years for a total of US $560,000.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Writing original draft, review, and editing: CE
Data curation, formal analysis, writing original draft, review, and editing: LR
Data curation and methodology: JL
Data curation: LDG, IT, NS, RS, ZS
Data curation, methodology, and formal analysis: SA
Supervision: JE, JY, DW
Project management: VT, TA
Funding, conceptualization, and visualization: MH
Project management, conceptualization, and visualization: LS, MT-K

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 579 KB - cardio_v10i1e82781_app1.pdf]

References

1. Kansra AR, Lakkunarajah S, Jay MS. Childhood and adolescent obesity: a review. Front Pediatr 2020;8(8):581461. [doi: 10.3389/fped.2020.581461] [Medline: 33511092]
2. Brauer M, Roth GA, Aravkin AY, et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet 2024 May;403(10440):2162-2203. [doi: 10.1016/S0140-6736(24)00933-4]

XSL•FO
RenderX

3. Valerio G, Di Bonito P, Calcaterra V, et al. Cardiometabolic risk in children and adolescents with obesity: a position paper of the Italian Society for Pediatric Endocrinology and Diabetology. Ital J Pediatr 2024 Oct 8;50(1):205. [doi: 10.1186/s13052-024-01767-x] [Medline: 39380079]
4. Skinner AC, Perrin EM, Moss LA, Skelton JA. Cardiometabolic risks and severity of obesity in children and young adults. N Engl J Med 2015 Oct;373(14):1307-1317. [doi: 10.1056/NEJMoa1502821] [Medline: 26422721]
5. Ogden CL, Emmerich SD, Stierman B, et al. Obesity among children and adolescents in NHANES August 2021-August 2023: an examination of race/Hispanic origin subgroup estimates. Pediatr Obes 2025 Oct;20(10):e70041. [doi: 10.1111/ijpo.70041] [Medline: 40915576]
6. Katz SF, Rodriguez F, Knowles JW. Health disparities in cardiometabolic risk among Black and Hispanic youth in the United States. Am J Prev Cardiol 2021 Jun;6:100175. [doi: 10.1016/j.ajpc.2021.100175] [Medline: 34327498]
7. Alberga AS, Sigal RJ, Goldfield G, Prud' homme D, Kenny GP. Overweight and obese teenagers: why is adolescence a critical period? Pediatr Obes 2012 Aug;7(4):261-273. [doi: 10.1111/j.2047-6310.2011.00046.x]
8. Ames ME, Leadbeater BJ, MacDonald SWS. Health behavior changes in adolescence and young adulthood: Implications for cardiometabolic risk. Health Psychol 2018 Feb;37(2):103-113. [doi: 10.1037/hea0000560] [Medline: 28967774]
9. Doom JR, Lumeng JC, Sturza J, Kaciroti N, Vazquez DM, Miller AL. Longitudinal associations between overweight/obesity and stress biology in low-income children. Int J Obes 2020 Mar;44(3):646-655. [doi: 10.1038/s41366-019-0447-4]
10. Griffith JM, Clark HM, Haraden DA, Young JF, Hankin BL. Affective development from middle childhood to late adolescence: trajectories of mean-level change in negative and positive affect. J Youth Adolesc 2021 Aug;50(8):1550-1563. [doi: 10.1007/s10964-021-01425-z] [Medline: 33791947]
11. Xiang AH, Martinez MP, Chow T, et al. Depression and anxiety among US children and young adults. JAMA Netw Open 2024 Oct 1;7(10):e2436906. [doi: 10.1001/jamanetworkopen.2024.36906] [Medline: 39352699]
12. Kokka I, Mourikis I, Bacopoulou F. Psychiatric disorders and obesity in childhood and adolescence—a systematic review of cross-sectional studies. Children (Basel) 2023 Feb 1;10(2):285. [doi: 10.3390/children10020285] [Medline: 36832413]
13. Lindberg L, Hagman E, Danielsson P, Marcus C, Persson M. Anxiety and depression in children and adolescents with obesity: a nationwide study in Sweden. BMC Med 2020 Feb 21;18(1):30. [doi: 10.1186/s12916-020-1498-z] [Medline: 32079538]
14. Roberts RE, Duong HT. Do anxiety disorders play a role in adolescent obesity? Ann Behav Med 2016 Aug;50(4):613-621. [doi: 10.1007/s12160-016-9786-8] [Medline: 26961207]
15. Leigh E, Clark DM. Understanding social anxiety disorder in adolescents and improving treatment outcomes: applying the cognitive model of Clark and Wells (1995). Clin Child Fam Psychol Rev 2018 Sep;21(3):388-414. [doi: 10.1007/s10567-018-0258-5] [Medline: 29654442]
16. O'Connor EA, Evans CV, Henninger M, Redmond N, Senger CA. Interventions for weight management in children and adolescents: updated evidence report and systematic review for the US Preventative Services Task Force. JAMA 2024 Jul 16;332(3):233-248. [doi: 10.1001/jama.2024.6739] [Medline: 38888913]
17. Adise S, Geier CF, Roberts NJ, White CN, Keller KL. Is brain response to food rewards related to overeating? A test of the reward surfeit model of overeating in children. Appetite 2018 Sep;128:167-179. [doi: 10.1016/j.appet.2018.06.014]
18. Favieri F, Marini A, Casagrande M. Emotional regulation and overeating behaviors in children and adolescents: a systematic review. Behav Sci (Basel) 2021 Jan 19;11(1):11. [doi: 10.3390/bs11010011] [Medline: 33477932]
19. Hussenoeder FS, Conrad I, Engel C, et al. Analyzing the link between anxiety and eating behavior as a potential pathway to eating-related health outcomes. Sci Rep 2021 Jul 19;11(1):14717. [doi: 10.1038/s41598-021-94279-1] [Medline: 34282257]
20. Goldschmidt AB, Smith KE, Lavender JM, Engel SG, Haedt-Matt A. Trait-level facets of impulsivity and momentary, naturalistic eating behavior in children and adolescents with overweight/obesity. J Psychiatr Res 2019 Mar;110:24-30. [doi: 10.1016/j.jpsychires.2018.12.018]
21. Ha OR, Lim SL. The role of emotion in eating behavior and decisions. Front Psychol 2023;14:1265074. [doi: 10.3389/fpsyg.2023.1265074] [Medline: 38130967]
22. Schaumberg K, Zerwas S, Goodman E, Yilmaz Z, Bulik CM, Micali N. Anxiety disorder symptoms at age 10 predict eating disorder symptoms and diagnoses in adolescence. J Child Psychol Psychiatry 2019 Jun;60(6):686-696. [doi: 10.1111/jcpp.12984] [Medline: 30353925]
23. Kerr-Gaffney J, Harrison A, Tchanturia K. Social anxiety in the eating disorders: a systematic review and meta-analysis. Psychol Med 2018 Nov;48(15):2477-2491. [doi: 10.1017/S0033291718000752] [Medline: 29631640]
24. Byrne ME, LeMay-Russell S, Tanofsky-Kraff M. Loss-of-control eating and obesity among children and adolescents. Curr Obes Rep 2019 Mar;8(1):33-42. [doi: 10.1007/s13679-019-0327-1] [Medline: 30701372]
25. Tanofsky-Kraff M, Shomaker LB, Olsen C, et al. A prospective study of pediatric loss of control eating and psychological outcomes. J Abnorm Psychol 2011 Feb;120(1):108-118. [doi: 10.1037/a0021406] [Medline: 21114355]
26. Tanofsky-Kraff M, Yanovski SZ, Schvey NA, Olsen CH, Gustafson J, Yanovski JA. A prospective study of loss of control eating for body weight gain in children at high risk for adult obesity. Int J Eat Disord 2009 Jan;42(1):26-30. [doi: 10.1002/eat.20580] [Medline: 18720473]

XSL•FO
RenderX

27. Ariño-Braña P, Zareba MR, Ibáñez Montolio M, Visser M, Picó-Pérez M. Influence of the HPA axis on anxiety-related processes: an RDoC overview considering their neural correlates. Curr Psychiatry Rep 2025 Oct;27(10):593-611. [doi: 10.1007/s11920-025-01633-5] [Medline: 40884631]
28. Bigalke JA, Durocher JJ, Greenlund IM, Keller-Ross M, Carter JR. Blood pressure and muscle sympathetic nerve activity are associated with trait anxiety in humans. Am J Physiol Heart Circ Physiol 2023 Apr 1;324(4):H494-H503. [doi: 10.1152/ajpheart.00026.2023] [Medline: 36800506]
29. Ravi M, Miller AH, Michopoulos V. The immunology of stress and the impact of inflammation on the brain and Behavior. BJPsych Adv 2021 May;27(Suppl 3):158-165. [doi: 10.1192/bja.2020.82] [Medline: 34055387]
30. Gianotti L, Belcastro S, D'Agnano S, Tassone F. The stress axis in obesity and diabetes mellitus: an update. Endocrines 2021;2(3):334-347. [doi: 10.3390/endocrines2030031]
31. Fontes MAP, Marins FR, Patel TA, et al. Neurogenic background for emotional stress-associated hypertension. Curr Hypertens Rep 2023 Jul;25(7):107-116. [doi: 10.1007/s11906-023-01235-7] [Medline: 37058193]
32. Wang Z, Luo Y, Zhang Y, et al. Heart rate variability in generalized anxiety disorder, major depressive disorder and panic disorder: a network meta-analysis and systematic review. J Affect Disord 2023 Jun;330:259-266. [doi: 10.1016/j.jad.2023.03.018]
33. Hassamal S. Chronic stress, neuroinflammation, and depression: an overview of pathophysiological mechanisms and emerging anti-inflammatories. Front Psychiatry 2023;14:1130989. [doi: 10.3389/fpsyt.2023.1130989] [Medline: 37252156]
34. Manigault AW, Shorey RC, Hamilton K, et al. Cognitive behavioral therapy, mindfulness, and cortisol habituation: a randomized controlled trial. Psychoneuroendocrinology 2019 Jun;104:276-285. [doi: 10.1016/j.psyneuen.2019.03.009] [Medline: 30917336]
35. Bhattacharya S, Goicoechea C, Heshmati S, Carpenter JK, Hofmann SG. Efficacy of cognitive behavioral therapy for anxiety-related disorders: a meta-analysis of recent literature. Curr Psychiatry Rep 2023 Jan;25(1):19-30. [doi: 10.1007/s11920-022-01402-8] [Medline: 36534317]
36. Paddon L, Kampman H. A new lens on improving physical health with psychological interventions: A systematic review. J Happiness Health 2023;3(2):49-92. [doi: 10.47602/johah.v3i2.46]
37. Banneyer KN, Bonin L, Price K, Goodman WK, Storch EA. Cognitive behavioral therapy for childhood anxiety disorders: a review of recent advances. Curr Psychiatry Rep 2018 Jul 28;20(8):65. [doi: 10.1007/s11920-018-0924-9] [Medline: 30056623]
38. Kendall PC. Child and Adolescent Therapy: Cognitive-Behavioral Procedures: Guilford Press; 2011.
39. Hilbert A, Petroff D, Neuhaus P, Schmidt R. Cognitive-behavioral therapy for adolescents with an age-adapted diagnosis of binge-eating disorder: a randomized clinical trial. Psychother Psychosom 2020;89(1):51-53. [doi: 10.1159/000503116] [Medline: 31533113]
40. Weissman MM, Markowitz JC, Klerman G. Comprehensive Guide to Interpersonal Psychotherapy: Basic Books; 2008.
41. Tanofsky-Kraff M, Shomaker LB, Young JF, Wilfley DE. Interpersonal psychotherapy for the prevention of excess weight gain and eating disorders: a brief case study. Psychotherapy (Chic) 2016 Jun;53(2):188-194. [doi: 10.1037/pst0000051] [Medline: 27267503]
42. Tanofsky-Kraff M, Shomaker LB, Wilfley DE, et al. Excess weight gain prevention in adolescents: three-year outcome following a randomized controlled trial. J Consult Clin Psychol 2017 Mar;85(3):218-227. [doi: 10.1037/ccp0000153] [Medline: 27808536]
43. Srivastava K, Chatterjee K, Prakash J, Yadav A, Chaudhury S. Comparative efficacy of cognitive behavior therapy and interpersonal therapy in the treatment of depression: a randomized controlled study. Ind Psychiatry J 2024;33(1):160-167. [doi: 10.4103/ipj.ipj_294_23] [Medline: 38853783]
44. Ekeblad A, Falkenström F, Andersson G, Vestberg R, Holmqvist R. Randomized trial of interpersonal psychotherapy and cognitive behavioral therapy for major depressive disorder in a community-based psychiatric outpatient clinic. Depress Anxiety 2016 Dec;33(12):1090-1098. [doi: 10.1002/da.22495] [Medline: 27029912]
45. Lucas LS, Borba BLI, de Azevedo BM, et al. Synchronous tele-interpersonal psychotherapy versus tele-cognitive behavioural therapy for adults: which works better? Results from a randomised clinical trial. Gen Psychiatr 2025;38(5):e102067. [doi: 10.1136/gpsych-2025-102067] [Medline: 41180136]
46. Zhou X, Hetrick SE, Cuijpers P, et al. Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: a systematic review and network meta-analysis. World Psychiatry 2015 Jun;14(2):207-222. [doi: 10.1002/wps.20217] [Medline: 26043339]
47. Dalle Grave R, Calugi S, Sartirana M, Sermattei S, Conti M. Enhanced cognitive behaviour therapy for adolescents with eating disorders: a systematic review of current status and future perspectives. IJEDO 2021;3:1. [doi: 10.32044/ijedo.2021.01]
48. Flynn L. The benefits and challenges of multisite studies: lessons learned. AACN Adv Crit Care 2009;20:388-391. [doi: 10.4037/15597768-2009-4013]
49. Orsolini L, Pompili S, Salvi V, Volpe U. A Systematic Review on TeleMental Health in Youth Mental Health: Focus on Anxiety, Depression and Obsessive-Compulsive Disorder. Medicina (Kaunas) 2021 Jul 31;57(8):793. [doi: 10.3390/medicina57080793] [Medline: 34440999]

XSL•FO
RenderX

50. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
51. Damian AJ, Stinchfield K, Kearney RT. Telehealth and beyond: promoting the mental well-being of children and adolescents during COVID. Front Pediatr 2022;10:793167. [doi: 10.3389/fped.2022.793167] [Medline: 35237538]
52. Wade SL, Gies LM, Fisher AP, et al. Telepsychotherapy with children and families: lessons gleaned from two decades of translational research. J Psychother Integr 2020 Jun;30(2):332-347. [doi: 10.1037/int0000215] [Medline: 34295136]
53. Franciosi EB, Tan AJ, Kassamali B, et al. The impact of telehealth implementation on underserved populations and no-show rates by medical specialty during the COVID-19 pandemic. Telemed J E Health 2021 Aug;27(8):874-880. [doi: 10.1089/tmj.2020.0525] [Medline: 33826411]
54. Rosenthal SR, Sonido PL, Tobin AP, Sammartino CJ, Noel JK. Breaking down barriers: young adult interest and use of telehealth for behavioral health services. R I Med J 2022;105(2):26-31 [FREE Full text]
55. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022;60(1):95-103. [doi: 10.1097/MLR.0000000000001664]
56. Spielberger CD, Edwards CD, Lushene R, Monturoi J, Platzek D. State-trait anxiety inventory for children: sampler set: manual, test booklet, scoring key. : Mind Garden; 1973 URL: https://www.scribd.com/document/926627389/STATE-TRAIT-ANXIETY-INVENTORY-FOR-CHILDREN-MANUAL [accessed 2026-09-07]
57. Cocks K, Torgerson DJ. Sample size calculations for pilot randomized trials: a confidence interval approach. J Clin Epidemiol 2013 Feb;66(2):197-201. [doi: 10.1016/j.jclinepi.2012.09.002] [Medline: 23195919]
58. Adler NE, Epel ES, Castellazzo G, Ickovics JR. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy, White women. Health Psychology 2000;19(6):586-592. [doi: 10.1037/0278-6133.19.6.586]
59. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016:i5239. [doi: 10.1136/bmj.i5239]
60. Rookes TA, Schrag A, Walters K, Armstrong M. Measures of fidelity of delivery and engagement in self-management interventions: a systematic review of measures. Clin Trials 2022 Dec;19(6):665-672. [doi: 10.1177/17407745221118555] [Medline: 36017707]
61. Hunsley J. Development of the Treatment Acceptability Questionnaire. J Psychopathol Behav Assess 1992 Mar;14(1):55-64. [doi: 10.1007/BF00960091]
62. Fairburn CG, Cooper Z. Binge eating: nature, assessment, and treatment. In: The Eating Disorder Examination, 12th edition: Guilford Press; 1993:317-360. [doi: 10.1016/0005-7967(94)90024-8]
63. Shain LM, Pao M, Tipton MV, et al. Comparing parent and child self-report measures of the State-Trait Anxiety Inventory in children and adolescents with a chronic health condition. J Clin Psychol Med Settings 2020 Mar;27(1):173-181. [doi: 10.1007/s10880-019-09631-5] [Medline: 31127422]
64. Beidel DC. Assessment of childhood social phobia: construct, convergent, and discriminative validity of the Social Phobia and Anxiety inventory for Children (SPA-C). Psychol Assess 1998;8(3):235-240. [doi: 10.1037/1040-3590.8.3.235]
65. Beidel DC, Turner SM, Hamlin K, Morris TL. The social phobia and anxiety inventory for children (SPAI-C): external and discriminative validity. Behav Ther 2000 ;31(1):75-87. [doi: 10.1016/S0005-7894(00)80005-2]
66. Tanofsky-Kraff M, Theim KR, Yanovski SZ, et al. Validation of the emotional eating scale adapted for use in children and adolescents (EES-C). Int J Eat Disord 2007 Apr;40(3):232-240. [doi: 10.1002/eat.20362] [Medline: 17262813]
67. Vannucci A, Tanofsky-Kraff M, Shomaker LB, et al. Construct validity of the emotional eating scale adapted for children and adolescents. Int J Obes (Lond) 2012 Jul;36(7):938-943. [doi: 10.1038/ijo.2011.225] [Medline: 22124451]
68. BodPod gs-x. COSMED. URL: https://www.cosmed.com/en/products/body-composition/bod-pod-gs-x [accessed 2026-08-31]
69. Child and teen BMI calculator. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/bmi/child-teen-calculator/index.html [accessed 2026-08-31]
70. McCrory MA, Gomez TD, Bernauer EM, Molé PA. Evaluation of a new air displacement plethysmograph for measuring human body composition. Med Sci Sports Exerc 1995 Dec;27(12):1686-1691. [doi: 10.1249/00005768-199512000-00016] [Medline: 8614326]
71. Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Med Res Methodol 2012 Dec 5;12:184. [doi: 10.1186/1471-2288-12-184] [Medline: 23216665]
72. Jong ST, Stevenson R, Winpenny EM, Corder K, van Sluijs EMF. Recruitment and retention into longitudinal health research from an adolescent perspective: a qualitative study. BMC Med Res Methodol 2023 Jan 16;23(1):16. [doi: 10.1186/s12874-022-01802-7] [Medline: 36647003]
73. Le Grange D, Eckhardt S, Dalle Grave R, et al. Enhanced cognitive-behavior therapy and family-based treatment for adolescents with an eating disorder: a non-randomized effectiveness trial. Psychol Med 2022 Oct;52(13):2520-2530. [doi: 10.1017/S0033291720004407] [Medline: 33267919]
74. Cassidy O, Eichen DM, Burke NL, et al. Engaging African American adolescents and stakeholders to adapt interpersonal psychotherapy for weight gain prevention. J Black Psychol 2018 Mar;44(2):128-161. [doi: 10.1177/0095798417747142] [Medline: 40969987]

XSL•FO
RenderX

75. Pena MA, Whitley MY, Sudarshan A, et al. Strategies to increase enrollment and retention in pediatric clinical research: a scoping review. Pediatrics 2025 Sep 1;156(Suppl 1):e2025070739O. [doi: 10.1542/peds.2025-070739O]
76. Hollon SD. Final Report: System for Rating Psychotherapy Audio Recordings: US Department of Health and Human Services; 1984.
77. Ottman KE, Kohrt BA, Pedersen GA, Schafer A. Use of role plays to assess therapist competency and its association with client outcomes in psychological interventions: a scoping review and competency research agenda. Behav Res Ther 2020 Jul;130:103531. [doi: 10.1016/j.brat.2019.103531] [Medline: 31902517]
78. Nye A, Delgadillo J, Barkham M. Efficacy of personalized psychological interventions: a systematic review and meta-analysis. J Consult Clin Psychol 2023 Jul;91(7):389-397. [doi: 10.1037/ccp0000820] [Medline: 37166831]
79. Simons AD, Marti CN, Rohde P, Lewis CC, Curry J, March J. Does homework "matter" in cognitive behavioral therapy for adolescent depression? J Cogn Psychother 2012;26(4):390-404. [doi: 10.1891/0889-8391.26.4.390]
80. Burriss KJ, Snead D. Middle school students' perceptions regarding motivation and effectiveness of homework. Sch Community J 2017;27(2) [FREE Full text]
81. Hudson JL, Kendall PC. Showing you can do it: homework in therapy for children and adolescents with anxiety disorders. J Clin Psychol 2002 May;58(5):525-534. [doi: 10.1002/jclp.10030]
82. Hjelmgren H, Nilsson A, Myrberg IH, Andersson N, Ygge BM, Nordlund B. Capillary blood sampling increases the risk of preanalytical errors in pediatric hospital care: observational clinical study. J Spec Pediatr Nurs 2021 Oct;26(4):e12337. [doi: 10.1111/jspn.12337] [Medline: 33960595]
83. Ibrahim JG, Chu H, Chen MH. Missing data in clinical studies: issues and methods. J Clin Oncol 2012 Sep 10;30(26):3297-3303. [doi: 10.1200/JCO.2011.38.7589] [Medline: 22649133]
84. Liverpool S, Mota CP, Sales CMD, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020;22(6):e16317. [doi: 10.2196/16317]
85. Keinonen K, Lappalainen P, Puolakanaho A, Kaipainen K, Lappalainen R, Kiuru N. Brief online ACT intervention to improve adolescents' well-being: effectiveness among adolescents with depressive symptoms during COVID-19. Child Youth Serv Rev 2024 Jun;161:107671. [doi: 10.1016/j.childyouth.2024.107671]
86. Vourdoumpa A, Paltoglou G, Manou M, et al. Improvement in symptoms of depression and anxiety and cardiometabolic risk factors in children and adolescents with overweight and obesity following the implementation of a multidisciplinary personalized lifestyle intervention program. Nutrients 2024 Oct 30;16(21):3710. [doi: 10.3390/nu16213710] [Medline: 39519542]
87. Scaini S, Belotti R, Ogliari A, Battaglia M. A comprehensive meta-analysis of cognitive-behavioral interventions for social anxiety disorder in children and adolescents. J Anxiety Disord 2016 Aug;42:105-112. [doi: 10.1016/j.janxdis.2016.05.008] [Medline: 27399932]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CSU: Colorado State University
HbA_{1c}: hemoglobin A_{1c}
IPT: interpersonal psychotherapy
IRB: institutional review board
STAI-C: State-Trait Anxiety Inventory for Children
USU: Uniformed Services University

Edited by A Coristine; submitted 21.Aug.2025; peer-reviewed by R Nooripour; revised version received 28.Jul.2026; accepted 10.Aug.2026; published 17.Sep.2026.

Please cite as:
Edwards C, Ranzenhofer L, Lavender J, Gulley LD, Aichele S, Thorstad I, Sanchez N, Schrag R, Sinkford Z, Emerick J, Thomas V, Arnold T, Young J, Wilfley D, Haigney M, Shomaker L, Tanofsky-Kraff M
Improving Cardiometabolic Health in Adolescent Girls With Elevated Anxiety: Web-Based Multisite, Pilot, and Feasibility Randomized Controlled Trial
JMIR Cardio 2026;10:e82781
URL: https://cardio.jmir.org/2026/1/e82781
doi:10.2196/82781

XSL•FO
RenderX

© Caitlin Edwards, LIsa Ranzenhofer, Jason Lavender, Lauren D Gulley, Stephen Aichele, Isabel Thorstad, Natalia Sanchez, Ruby Schrag, Zoe Sinkford, Jill Emerick, Victoria Thomas, Thomas Arnold, Jami Young, Denise Wilfley, Mark Haigney, Lauren Shomaker, Marian Tanofsky-Kraff. Originally published in JMIR Cardio (https://cardio.jmir.org), 17.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association Between Type D Personality and Cardiovascular Disease History: Cross-Sectional Study

Keren Grinberg, RN, PhD; Yael Sela, RN, PhD

Nursing Sciences Department, Faculty of Social and Community Sciences, Ruppin Academic Center, Emek Hefer, Israel

Corresponding Author:
Keren Grinberg, RN, PhD
Nursing Sciences Department, Faculty of Social and Community Sciences, Ruppin Academic Center, Emek Hefer, Israel

Abstract

Background: Type D personality, characterized by high negative affectivity and social inhibition, has been linked to poorer mental health and heightened risk for adverse cardiovascular outcomes. Although previous studies have examined associations between type D personality, psychological distress, and cardiovascular disease (CVD), many have assessed these factors independently, relied on clinical samples, or overlooked the simultaneous assessment of psychological distress and CVD history. Consequently, less is known about how type D traits relate to emotional distress and CVD history within the general population. Understanding these relationships may support early identification of at-risk individuals and strengthen the integration of psychological screening into cardiovascular care.

Objective: This study aimed to (1) examine associations between type D personality, emotional distress (depression, anxiety, and stress), and self-reported CVD history; (2) compare distress levels among participants with and without CVD history; and (3) determine whether type D personality predicts emotional distress independent of demographic factors and CVD history.

Methods: A cross-sectional online survey was completed by 146 adults aged 30 to 85 years, recruited through convenience and snowball sampling on social media. Type D personality was assessed using the Type D Scale–14, and emotional distress was measured using the Depression Anxiety and Stress Scale-21 items. CVD history was captured through a single self-report question regarding prior diagnosis of a cardiovascular condition. Descriptive statistics characterized the sample. Two-tailed independent samples *t* tests compared distress between individuals with and without type D personality and between participants with and without CVD history. Pearson correlation coefficients examined associations among key variables. Hierarchical multiple regression assessed whether type D personality predicted emotional distress beyond age, gender, education, and CVD history.

Results: Of the 146 participants, 40 (27.4%) reported a history of CVD and 62 (42.5%) met criteria for type D personality. Individuals with type D personality exhibited significantly higher depression, anxiety, and stress levels than non–type D participants (all $P<.001$). Participants with CVD history also reported greater distress compared with those without CVD history. Hierarchical regression analyses showed that type D personality remained a strong independent predictor of emotional distress ($\beta=.46$; $P<.001$) after adjusting for demographics and CVD history. CVD history made an additional but smaller contribution to distress ($\beta=.18$; $P=.008$). These findings highlight the cumulative influence of personality traits and cardiovascular background on psychological well-being.

Conclusions: Type D personality traits have been associated with higher levels of psychological distress and with a greater likelihood of self-reported CVD history in the general population. Type D personality remained a significant predictor of distress after accounting for demographic factors and cardiovascular history, underscoring its potential role in early psychological risk identification. Incorporating brief personality and mental health screening into cardiovascular assessment may support more comprehensive care.

(JMIR Cardio 2026;10:e79159) doi:10.2196/79159

KEYWORDS

type D personality; cardiovascular disease; psychological distress; depression; anxiety; stress

Introduction

Background and Rationale

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, resulting from a complex interplay of genetic, physiological, behavioral, and psychosocial factors [1,2]. While traditional risk factors such as age, hypertension, smoking, and dyslipidemia are well established, there is growing recognition of the contribution of psychological characteristics to cardiovascular risk and prognosis [3,4]. Among these, type D ("distressed") personality has drawn attention due to its association with poor mental health and adverse cardiac outcomes [5].

XSL•FO
RenderX

Type D personality is defined by the joint presence of negative affectivity (NA), the tendency to experience negative emotions, and social inhibition (SI), the tendency to suppress emotional expression and avoid social interactions [6]. Individuals with this personality profile are more likely to report symptoms of depression, anxiety, and stress and to experience low social support, poor quality of life, and unfavorable health behaviors [7,8]. This personality pattern is considered relatively stable over time and is found in both clinical and nonclinical populations [9]. Habibović et al [10] found that among patients with CVD, type D personality might be associated with lower social engagement, which could, in turn, partly explain its association with adverse health outcomes.

Emerging evidence suggests that type D personality is associated with an increased risk of cardiovascular morbidity and mortality, independent of traditional risk factors. Proposed mechanisms include heightened physiological stress responses (eg, hypothalamic-pituitary-adrenal axis dysregulation and inflammation), maladaptive coping strategies, and delayed health-seeking behavior [11,12]. Moreover, type D personality may interact synergistically with emotional distress to compound cardiovascular risk [13,14].

Despite increasing interest in this construct, several gaps remain. First, many studies have focused on patients with known cardiac diagnoses, limiting generalizability to broader populations. Second, type D personality and emotional distress (eg, depression, anxiety, and stress) are often examined in isolation, rather than concurrently, despite their theoretical and empirical overlap. Third, there is a lack of studies conducted in non-Western or diverse sociocultural contexts, such as Israel, where demographic and psychosocial profiles may differ.

This study addresses these gaps by simultaneously examining type D personality, psychological distress, and self-reported CVD history in a community sample. Unlike prior work focusing exclusively on clinical end points, our goal is to characterize the psychological profile of adults with and without cardiac history to assess whether type D personality is associated with greater emotional distress and cardiovascular vulnerability.

Objectives and Hypotheses

This study has two primary aims: (1) to examine whether individuals with type D personality report higher levels of depression, anxiety, and stress compared to non–type D individuals; and (2) to assess whether type D personality is an independent predictor of psychological distress after controlling for age, gender, education, and cardiovascular history.

On the basis of prior literature and theoretical models, we hypothesize the following:

- Hypothesis 1—individuals with type D personality will report significantly higher emotional distress (depression, anxiety, and stress) than those without type D traits.
- Hypothesis 2—type D personality will significantly predict psychological distress after adjusting for demographic and medical variables.

By testing these hypotheses, this study aimed to contribute to a more nuanced understanding of how personality traits relate to psychological well-being and cardiovascular history. Such knowledge may help inform early screening and targeted interventions in cardiac care settings.

Methods

Study Design

A cross-sectional, quantitative study was conducted to assess associations between type D personality, psychological distress, and self-reported cardiovascular history in a nonclinical adult sample. The study was part of a larger survey-based investigation examining personality and health factors.

Participants and Recruitment

A total of 146 participants were recruited via convenience and snowball sampling using targeted advertisements on Facebook and WhatsApp (Meta Platforms, Inc) between January and April 2022. Inclusion criteria included being aged 30 years or older, fluency in Hebrew, and access to the internet. Participants completed an anonymous online questionnaire hosted on Google Forms (Multimedia Appendix 1). Sensitivity analysis confirmed that the sample size provided sufficient statistical power for the logistic and linear regression models used.

Ethical Considerations

The study was approved by the Institutional Review Board of Ruppin Academic Center (approval code 251-L/22). Informed consent was obtained electronically from all participants before participation. All responses were anonymous. Participants received no monetary compensation. All data were deidentified before analysis.

Measures

Demographics and Health History

A background questionnaire collected data on age, gender, marital status, number of children, education level, religious identification, and self-reported cardiovascular history. CVD history was determined via the following question: “Have you ever been diagnosed with a cardiovascular condition or experienced a cardiac event (Yes/No)”?

Type D Personality

The Type D Scale–14 [8] was used to measure type D personality. The scale includes 14 items rated on a 5-point Likert scale, yielding 2 subscales: NA and SI. Participants scoring 10 or greater on both subscales were classified as type D. The scale has demonstrated good reliability in Hebrew-speaking populations (α>.80).

Psychological Distress

The Depression Anxiety Stress Scale (DASS)-21 items [13] was used to assess emotional distress. The 21 items measure depression, anxiety, and stress over the previous week using a 4-point Likert scale. Subscale scores were calculated and multiplied by 2 to match standard interpretations. Internal consistency was high (α=.87-.91).

XSL•FO
RenderX

Data Screening and Statistical Analyses

Data were screened for completeness, normality, and outliers. All variables met the assumptions for parametric testing. Descriptive statistics were computed for all variables.

Independent samples *t* tests and chi-square tests compared psychological outcomes and type D classification across demographic groups and CVD history. Pearson correlations examined associations among continuous variables.

Hierarchical linear regression was used to evaluate whether type D personality predicted psychological distress (total DASS score), after controlling for age, gender, education, and CVD history. All analyses were conducted using SPSS (version 28; IBM Corp). Two-tailed independent samples *t* tests with $P<.05$ were considered significant. Effect sizes (ie, Cohen *d*, η^2, and R^2) were reported.

Results

Sample Characteristics

Of the 146 participants, 40 (27.4%) reported a history of CVD. The sample was predominantly Jewish (140/146, 95.9%), secular (102/146, 70%), and highly educated (116/146, 80% held postsecondary degrees). Table 1 presents detailed demographic and psychological characteristics. Participants were categorized into 6 age groups: 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 to 85 years. The age distribution indicated that the majority of participants (95/146, 65%) were aged 50 to 70 years, while an additional 20% (29/146) were aged 70 to 85 years. Younger participants were less represented, with 5% (8/146) aged 30 to 39 years and 10% (15/146) aged 40 to 49 years. In total, 66% (97/146) of participants were female.

Table . Descriptive statistics and internal consistency of psychological measures among Israeli adults (N=146).

Variable	Values, mean (SD)	Cronbach α
Negative affectivity	11.4 (6.7)	0.89
Social inhibition	13.2 (6.1)	0.92
Depression	14.6 (5.8)	0.87
Anxiety	12.7 (5.4)	0.89
Stress	15.9 (6.5)	0.91
Distress (total)	43.2 (14.9)	0.94

Prevalence of Type D Personality

On the basis of the Type D Scale–14, 42% (62/146) of participants met the criteria for type D personality . Type D participants were more likely to report a history of CVD (54/146, 36.9%) than non–type D participants (29/146, 20.2%; $\chi^2_1=4.5$; $P=.03$).

Group Differences in Psychological Distress

Participants with type D personality scored significantly higher on depression (mean 19.2, SD 5.0) than non–type D participants (mean 11.3, SD 4.5; $P<.001$; Cohen $d=1.72$), indicating a large effect size. Similarly, anxiety (mean 16.5 vs 9.7, SD 4.6 vs 3.9; $P<.001$; Cohen $d=1.63$) and stress (mean 21.0 vs 12.1, SD 5.3 vs 4.8; $P<.001$; Cohen $d=1.76$) were markedly higher in type D participants, reflecting strong group differences. Psychological distress differed between the study groups, as shown in Table 2.

Table . Group differences in depression, anxiety, and stress between participants with and without type D personality, based on summed Depression Anxiety Stress Scale–21 items subscale scores (range 0 - 21; N=146).

Variable	Type D (n=62), mean (SD)	Non–type D (n=84), mean (SD)	*t* value (*df*)	*P* value
Depression	19.2 (5.0)	11.3 (4.5)	8.12 (144)	<.001
Anxiety	16.5 (4.6)	9.7 (3.9)	7.89 (144)	<.001
Stress	21.0 (5.3)	12.1 (4.8)	8.48 (144)	<.001

Associations With Cardiovascular History

Participants with CVD history had higher mean type D scores (mean 1.42, SD 0.61) and higher levels of psychological distress than those without CVD. Depression, anxiety, and stress were significantly elevated among those with CVD (all $P<.05$).

Regression Analysis

A 2-step hierarchical regression evaluated predictors of psychological distress. Step 1 included demographics and CVD history and explained 14.6% of the variance ($F_{4,141}=6.05$; $P<.001$). Step 2 added type D personality and accounted for an additional 15.1% of variance ($\Delta F=29.64$; $P<.001$), with the final model explaining 29.7% of total variance ($F_{5,140}=14.87$; $P<.001$).

Type D personality was the strongest predictor ($\beta=.46$; $P<.001$), followed by CVD history ($\beta=.18$; $P=.008$). Age, gender, and education were not significant.

XSL•FO
RenderX

Discussion

Principal Findings

This study found that type D personality was strongly associated with elevated depression, anxiety, and stress in a community-based Israeli sample, with effects persisting after adjustment for age, gender, education, and CVD history. Participants with CVD also had higher type D scores and distress, although personality traits showed a stronger effect. The prevalence of type D (62/146, 42%) exceeded global community estimates, potentially reflecting cultural or contextual factors. This unusually high prevalence may reflect specific cultural, societal, or psychosocial characteristics within the Israeli context, such as collectivist norms, social stressors, or health awareness, highlighting the importance of considering local sociocultural factors when interpreting type D personality prevalence and its psychological correlates. These results align with epidemiologic evidence showing shifts in cardiovascular determinants and the growing role of psychosocial risk factors [15-17], as well as updates highlighting stress pathways in CVD pathophysiology [16]. Depression levels in participants with type D personality were in the moderate range (mean 19.2, SD 5.0), consistent with literature linking depressive symptoms to increased cardiac risk via behavioral, autonomic, inflammatory, and platelet activation mechanisms [18,19]. Such clustering of distress in individuals with type D personality mirrors psychocardiology findings that integrate psychological traits with cardiovascular biology and epidemiology [20].

By jointly examining type D personality, psychological distress, and CVD history, this study extends previous work focused mainly on clinical cardiac populations, demonstrating similar associations in nonclinical adults. Mechanistically, high NA and SI may limit social support, encourage maladaptive coping, and promote biological dysregulation, all of which can contribute to cardiovascular vulnerability [16,18]. The findings suggest potential value in community and primary care screening for type D traits and distress, in line with clinical reviews emphasizing early psychosocial intervention to improve cardiac outcomes [19,21]. Although the cross-sectional design limits causal inference, and the sample's demographic profile may affect generalizability, these results support a multifactorial model where stable personality traits, affective states, and cardiovascular history interact within the broader epidemiologic trends in CVD [15-17,20].

It should be noted that because type D personality includes NA and the outcome measure (DASS-21) assesses depression, anxiety, and stress—all facets of negative affect—our analysis may partly reflect conceptual overlap rather than an independent predictive effect, highlighting a tautological limitation that warrants careful interpretation.

Limitations

Several limitations warrant discussion. First, the use of self-reported CVD history introduces potential misclassification and recall bias. Although we included examples in the survey question, objective verification (eg, medical records) was not feasible. Additionally, the use of a single self-report item for CVD history limits granularity and may affect associations with type D personality and psychological distress.

Second, the cross-sectional design precludes causal inference. While we propose that type D traits may precede emotional distress, it is also plausible that chronic distress shapes personality expression (reverse causality). Future longitudinal research is needed to disentangle these pathways. In addition, the convenience and snowball sampling may have introduced self-selection bias, as individuals with greater psychological concerns or distress may have been more likely to participate, possibly inflating type D prevalence and its association with distress.

Third, the sample was predominantly female, Jewish (140/146, 95.9%), highly educated, and secular. This demographic specificity, along with the nonrandom recruitment strategy, limits the generalizability of the findings. The prevalence of type D personality in our sample (62/146, 42%) was higher than typically reported in community-based studies, which may reflect characteristics of the recruited population and further limits external validity.

Fourth, the overlap between type D (particularly the NA subscale) and general negative affect raises concerns about multicollinearity in statistical models. Future studies should consider examining subscale interactions or partial correlations to clarify unique contributions.

Fifth, our analyses did not account for several established cardiovascular risk factors, including behavioral factors (eg, smoking, physical activity, and diet), biomedical factors (eg, BMI, diabetes, and dyslipidemia), and broader socioeconomic variables (eg, income and employment status). These unmeasured variables may confound the observed associations between type D personality, psychological distress, and CVD history and should be considered in future research.

Future research should use longitudinal and multimethod designs to evaluate the causal relationships among personality traits, emotional distress, and cardiovascular outcomes. Clinical studies involving medically verified diagnoses, as well as diverse cultural and demographic samples, are needed to enhance external validity. In addition, the development and validation of brief screening tools for type D personality in primary care and cardiology settings could facilitate early identification. Intervention studies should examine whether reducing emotional distress in individuals with type D personality leads to improvements in cardiovascular outcomes.

Recommendations

Future studies should further explore the clinical implications of incorporating psychological screening, such as type D personality, into cardiovascular care. Additional research is warranted to evaluate how integrating psychological assessment into routine practice may contribute to more personalized prevention and intervention strategies.

Conclusions

This study demonstrates that type D personality is significantly associated with elevated psychological distress and a history of CVD, even within a community sample. These findings

XSL•FO
RenderX

highlight the importance of considering psychological factors when examining cardiovascular health outcomes.

Funding

This research received no external funding.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: KG
Data collection: KG, YS
Formal analysis: KG
Methodology: KG
Supervision: KG
Writing – original draft: KG, YS
Writing – review and editing: KG, YS
All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey of the study (in Hebrew).
[DOCX File, 25 KB - cardio_v10i1e79159_app1.docx]

References

1. Cardiovascular diseases. World Health Organization. 2025. URL: https://www.who.int/health-topics/cardiovascular-diseases [accessed 2025-04-20]
2. Cruz-Martínez RR, Wentzel J, Bente BE, Sanderman R, van Gemert-Pijnen JE. Toward the value sensitive design of eHealth technologies to support self-management of cardiovascular diseases: content analysis. JMIR Cardio 2021 Dec 1;5(2):e31985. [doi: 10.2196/31985] [Medline: 34855608]
3. Grande G, Romppel M, Barth J. Association between type D personality and prognosis in patients with cardiovascular diseases: a systematic review and meta-analysis. Ann Behav Med 2012 Jun;43(3):299-310. [doi: 10.1007/s12160-011-9339-0] [Medline: 22237826]
4. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts): developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016 Aug 1;37(29):2315-2381. [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
5. Chauvet-Gelinier JC, Bonin B. Stress, anxiety and depression in heart disease patients: a major challenge for cardiac rehabilitation. Ann Phys Rehabil Med 2017 Jan;60(1):6-12. [doi: 10.1016/j.rehab.2016.09.002] [Medline: 27771272]
6. Sajobi TT, Wang M, Awosoga O, et al. Trajectories of health-related quality of life in coronary artery disease. Circ Cardiovasc Qual Outcomes 2018 Mar;11(3):e003661. [doi: 10.1161/CIRCOUTCOMES.117.003661] [Medline: 29545392]
7. Denollet J. Type D personality. A potential risk factor refined. J Psychosom Res 2000 Oct;49(4):255-266. [doi: 10.1016/s0022-3999(00)00177-x] [Medline: 11119782]
8. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. Psychosom Med 2005;67(1):89-97. [doi: 10.1097/01.psy.0000149256.81953.49] [Medline: 15673629]
9. Mols F, Denollet J. Type D personality in the general population: a systematic review of health status, mechanisms of disease, and work-related problems. Health Qual Life Outcomes 2010 Jan 23;8:9. [doi: 10.1186/1477-7525-8-9] [Medline: 20096129]
10. Habibović M, Gavidia G, Broers E, et al. Type D personality and global positioning system tracked social behavior in patients with cardiovascular disease. Health Psychol 2020 Aug;39(8):711-720. [doi: 10.1037/hea0000823] [Medline: 32297772]
11. Henein MY, Vancheri S, Longo G, Vancheri F. The impact of mental stress on cardiovascular health-part II. J Clin Med 2022 Jul 28;11(15):4405. [doi: 10.3390/jcm11154405] [Medline: 35956022]

XSL•FO
RenderX

12. Lear SA, McKee M, Hystad P, et al. Social factors, health policy, and environment: implications for cardiovascular disease across the globe. Eur Heart J 2025 Aug 8;46(30):2959-2973. [doi: 10.1093/eurheartj/ehaf212] [Medline: 40259769]
13. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
14. Kupper N, Denollet J. Type D personality as a risk factor in coronary heart disease: a review of current evidence. Curr Cardiol Rep 2018 Sep 12;20(11):104. [doi: 10.1007/s11886-018-1048-x] [Medline: 30209683]
15. Lv H, et al. Impact of type D personality on major adverse cardiac events. J Psychosom Res 2020;136:110192. [doi: 10.1016/j.jpsychores.2020.110192]
16. Mensah GA, Wei GS, Sorlie PD, et al. Decline in cardiovascular mortality: possible causes and implications. Circ Res 2017 Jan 20;120(2):366-380. [doi: 10.1161/CIRCRESAHA.116.309115] [Medline: 28104770]
17. Vaccarino V, Bremner JD. Stress and cardiovascular disease: an update. Nat Rev Cardiol 2024 Sep;21(9):603-616. [doi: 10.1038/s41569-024-01024-y] [Medline: 38698183]
18. McAloon CJ, Boylan LM, Hamborg T, et al. The changing face of cardiovascular disease 2000-2012: an analysis of the world health organisation global health estimates data. Int J Cardiol 2016 Dec 1;224:256-264. [doi: 10.1016/j.ijcard.2016.09.026] [Medline: 27664572]
19. Carney RM, Freedland KE. Depression and coronary heart disease. Nat Rev Cardiol 2017 Mar;14(3):145-155. [doi: 10.1038/nrcardio.2016.181] [Medline: 27853162]
20. Hare DL, Toukhsati SR, Johansson P, Jaarsma T. Depression and cardiovascular disease: a clinical review. Eur Heart J 2014 Jun 1;35(21):1365-1372. [doi: 10.1093/eurheartj/eht462] [Medline: 24282187]
21. You Y, Shou X, Zhang X, et al. Psycho-cardiological disease: a bibliometric review from 2001 to 2021. Front Cardiovasc Med 2022;9:890329. [doi: 10.3389/fcvm.2022.890329] [Medline: 35571163]

Abbreviations

CVD: cardiovascular disease
DASS-21: Depression Anxiety Stress Scales
NA: negative affectivity
SI: social inhibition

Edited by A Coristine; submitted 16.Jun.2025; peer-reviewed by M Whited, R Nissanholtz-Gannot, SB Guo; revised version received 23.Dec.2025; accepted 23.Dec.2025; published 10.Mar.2026.

Please cite as:
Grinberg K, Sela Y
Association Between Type D Personality and Cardiovascular Disease History: Cross-Sectional Study
JMIR Cardio 2026;10:e79159
URL: https://cardio.jmir.org/2026/1/e79159
doi:10.2196/79159

© Keren Grinberg, Yael Sela. Originally published in JMIR Cardio (https://cardio.jmir.org), 10.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Group-Based Multitrajectory Modeling Analysis of Joint Trajectories of Blood Pressure and Heart Rate and Their Association With Relative Change in Relative Wall Thickness in Patients With Hypertension: Community-Based Longitudinal Cohort Study

Qi Pan[1], MPH; Nanxiang Zhang[1], PhD; Shuo Yang[1], PhD; Chonglong Ding[1], MMSc; Jianan Yin[1], MPH; Yi Fei[1], BSc; Xinyan Zou[1], BMSc; Yongjun Zheng[2], BSc; Xiang Huang[2], PhD; Jiezhen Feng[2], BSc; Hai Lin[2], BSc; Jinxin Zhang[1], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, No. 74, Zhongshan 2nd Road, Yuexiu District, Guangzhou, China
[2]Sanxiang Community Health Service Center of Zhongshan, Town Goverment of Zhongshan, Zhongshan, China

Corresponding Author:
Jinxin Zhang, PhD
Department of Medical Statistics, School of Public Health, Sun Yat-sen University, No. 74, Zhongshan 2nd Road, Yuexiu District, Guangzhou, China

Abstract

Background: Ambulatory blood pressure monitoring objectively assesses circadian hemodynamic rhythms. However, the joint trajectories of longitudinal blood pressure (BP) and heart rate (HR), and their association with early hypertensive cardiac remodeling, remain incompletely characterized.

Objective: This study aimed to identify joint trajectories of BP and HR using group-based multitrajectory modeling and examine their associations with relative changes in relative wall thickness (ΔRWT).

Methods: This community-based longitudinal cohort study included 213 patients with hypertension, with data collected between November 2022 and January 2025. Participants underwent 4 ambulatory BP monitoring sessions over 6 months, and echocardiography was measured at 3 months and 9 months. Group-based multitrajectory modeling was used to identify joint trajectories of 24-hour mean systolic blood pressure, diastolic blood pressure, and HR. A full-model multiple linear regression estimated the associations between trajectory groups and ΔRWT. Sensitivity analyses were conducted by adjusting for age and sex, and by using complete-case data.

Results: Four joint trajectory groups were identified: group 1 (lowest BP and HR, n=59), group 2 (normal BP, high HR, n=60), group 3 (stage 1 hypertension range, normal HR, n=54), and group 4 (stage 2 hypertension range, high HR, n=40). In the primary adjusted multiple linear regression model, compared with group 4, group 2 showed a significant inverse association with changes in ΔRWT (*b*=−0.059, 95% CI −0.112 to −0.007; *P*=.03). Antidiabetic medication use was significantly associated with a reduction in ΔRWT (*b*=−0.067, 95% CI −0.116 to −0.018; *P*=.008). Waking up before 7 AM was positively associated with an increase in ΔRWT (*b*=0.045, 95% CI 0.007-0.083; *P*=.02). Notably, the association for group 2 attenuated to marginal significance (*P*=.06) in the complete-case analysis, primarily due to the reduction in statistical power.

Conclusions: Hypertensive cardiac remodeling exhibits significant heterogeneity across distinct joint systolic blood pressure, diastolic blood pressure, and HR trajectories. Stable BP control is associated with a potential attenuation of early cardiac remodeling even in the presence of an elevated HR; however, this beneficial association attenuated to marginal significance in the complete-case analysis. Concurrent antidiabetic medication use and waking after 7 AM showed a beneficial association with the relative change in ΔRWT.

(JMIR Cardio 2026;10:e87091) doi:10.2196/87091

KEYWORDS

hypertension; ambulatory blood pressure monitoring; ventricular wall thickness; cardiac remodeling; group-based multitrajectory modeling

XSL•FO
RenderX

Introduction

Hypertension is a chronic cardiovascular disease (CVD). It involves persistent high arterial blood pressure (BP) and is a top cause of global cardiovascular illness and early death [1,2]. High systolic blood pressure (SBP) was one of the top 3 risk factors for poor health and early death worldwide in 2021, along with smoking and high blood sugar [3]. Its prevalence doubled from 650 million (1990) to 1.3 billion adults (2019) [4].

Ambulatory blood pressure monitoring (ABPM) can objectively and comprehensively assess circadian rhythms and occult hypertension [5,6], and several community-based studies have shown that ambulatory BP is superior to office BP in determining cardiovascular risk [7-9]. Poor BP control raises cardiovascular event and death risks significantly [10]. Hypertension can lead to progressive deterioration of left ventricular (LV) geometry, an important predictor of cardiovascular events [11-13]. Four LV geometric patterns are defined using left ventricular mass index (LVMI) and relative wall thickness (RWT): concentric hypertrophy (CH), eccentric hypertrophy, concentric remodeling, and normal geometry [14]. Left ventricular mass exhibits a continuous association with cardiovascular risk [15]. CH, defined by an elevated LVMI accompanied by RWT, has been shown to be a strong predictor of future cardiovascular events in patients with hypertension [12,16]. RWT serves as a central indicator to distinguish CH from eccentric hypertrophy and directly determines the pathological features of CH [17-19]. Although global wall thickness shows superior discriminatory power in patients with early-stage normal left ventricular mass [20], prognostic studies focusing specifically on RWT remain limited. Given the important role of RWT in LV remodeling patterns and its predictive value for cardiovascular prognosis, this study chooses it as an outcome indicator representing LV geometry.

However, traditional clinical practice relies on a single office BP measurement, which makes it difficult to capture the dynamic changes of BP (ie, BP trajectory) [21]. A single measurement is susceptible to stochastic fluctuation, which may lead to erroneous risk stratification [22,23]. Group-based trajectory modeling identifies the number of subgroups with different developmental trends from longitudinal data and fits the developmental trajectories of different subgroups [24,25]. Extending this framework, group-based multitrajectory modeling (GBMTM) allows researchers to analyze associations among several relevant outcomes over time and across subgroups. In cardiovascular epidemiology, group-based trajectory modeling reveals adverse change patterns in glucose and BP trajectories that predict coronary and cerebrovascular events [26,27]. Recent studies have confirmed that BP trajectories are better predictors of cardiovascular prognosis than a single measurement [28]. However, the relationship between the joint trajectories of BP and heart rate (HR) and cardiac structure in patients with hypertension remains unclear. Previous studies have analyzed SBP, diastolic blood pressure (DBP), and HR as independent variables, overlooking their physiological synergy. However, BP and HR are coregulated by the autonomic nervous system and are often simultaneously influenced by sedatives, analgesics, and antihypertensive drugs [29,30]. Patients with hypertension often exhibit increased resting HR, with more pronounced elevation in those with moderate to severe hypertension [31,32].

Therefore, this study aimed to identify distinct joint trajectories of 24-hour mean SBP, DBP, and HR among community-based patients with hypertension using GBMTM, and to investigate the associations between these trajectory groups and ΔRWT, a sensitive marker of early hypertensive cardiac remodeling.

Methods

Study Participants

The study cohort (Dynamic ECG & Blood Pressure for Hypertension at Sanxiang) consisted of managed care hypertension cases within the primary care community of Zhongshan City, Guangdong Province, China, from November 1, 2022, to January 1, 2025 [33]. The inclusion criteria for this study were based on meeting the diagnostic criteria for hypertension in the Chinese Hypertension Prevention and Treatment Guidelines (2024 Revised Edition) [34]: SBP ≥140 mm Hg and/or DBP ≥90 mm Hg measured on 3 non–consecutive days within a week. Patients with hypertension who were included in the management of the National Basic Public Health Service Programme were included in the study. Inclusion criteria included baseline age ≥35 years; clear consciousness, unimpeded communication, and no cognitive impairment; and completion of 4 ABPM sessions (at baseline, 3 months, 6 months, and 9 months after baseline; with 3-month intervals). Exclusion criteria were patients with secondary hypertension (such as renovascular hypertension, endocrine-induced hypertension, etc); individuals with gestational hypertension; those complicated with severe acute or chronic diseases (eg, malignant tumors, end-stage renal disease, and acute decompensated heart failure) or severe physical disabilities; those with incomplete baseline data (eg, missing key variables) or inability to cooperate with follow-up; and those who withdrew voluntarily or were lost to follow-up.

Data Collection

Structured Questionnaire

Through reviewing the literature [35-37], a structured, face-to-face questionnaire was administered to collect comprehensive baseline data. This instrument captured sociodemographic characteristics (age, sex, residency, education, and marital status), physical metrics (BMI), lifestyle habits (smoking and alcohol consumption), and current medication use. To specifically evaluate hypertension management, the questionnaire also assessed recent dietary patterns (primarily evaluating the structural balance between meat-based and plant-based diets), history of diabetes (verified via patient self-reporting combined with community electronic health records), and self-reported average morning wake-up time.

For statistical analysis, self-reported morning wake-up times were dichotomized at 7 AM. This specific threshold is grounded in established chronobiological and cardiovascular physiological evidence [38]. During this early morning window, the morning BP surge and cortisol awakening response reach their peak, which corresponds directly to the well-documented surge in

XSL•FO
RenderX

major cardiovascular events occurring between 6 AM and 12 PM [39,40]. Physiologically, awakening before 7 AM triggers a premature shift toward sympathetic nervous system dominance and amplifies neuroendocrine stress responses [41,42]; both mechanisms are intrinsically linked to elevated vascular resistance and subsequent adverse cardiac remodeling.

Sleep Status Rating Scale

The Self-Rating Scale of Sleep [43] was used to assess the sleep status of the patients. There are 10 items in the scale, and each item is rated on a 5-point scale (1-5), with higher scores indicating more serious sleep problems. Li [43] initially proposed that a total Self-Rating Scale of Sleep score of ≥23 indicates sleep problems.

Ambulatory Blood Pressure Monitoring

After completion of baseline data collection, patients with hypertension who completed the baseline assessment received regular follow-up visits. At each visit, we used a certified 24-hour ABPM (Nalon ABP01, Nalon Health Technology Co, Ltd) for 24-hour continuous monitoring (valid readings ≥80%, recorded every 30/60 minutes during daytime or nighttime, respectively), and systematically collected 24-hour mean SBP, DBP, and HR data to comprehensively assess the patients' BP status.

Echocardiography

Echocardiography was performed at 3 months after baseline and 9 months after baseline using a Philips EPIQ CVx ultrasound system with an S5-1 probe (Royal Philips) to objectively and comprehensively obtain the patient's cardiac-related parameters. RWT was calculated according to the American Society of Echocardiography recommendations [44] as follows: RWT= (2×posterior wall thickness at end-diastole)/left ventricular internal dimension at end-diastole (LVIDd). To strictly control measurement bias, echocardiographic examinations for the same patient were independently performed by a doctor, while the subsequent entry of echocardiographic parameters was handled by 2 separate investigators. Ultimately, all collected data underwent rigorous range and logic checks to ensure database completeness and accuracy.

Medication Exposure

Medication exposure during the 9-month follow-up was tracked via prescription refill records, categorizing the predominant regimen into drug classes including angiotensin receptor blockers (ARBs), calcium channel blockers (CCBs), β-blockers, and combination therapies. To account for potential treatment adjustments during this period, each patient's predominant medication regimen was defined as the specific drug class or combination prescribed most frequently across all outpatient visits.

Quality Control

To ensure data quality for the managed patients with hypertension, multiple strategies were implemented. All follow-up staff were professionally trained and certified. They strictly adhered to standardized operating procedures during data collection to ensure the consistency and objectivity of raw data.

During data cleaning and validation, abnormal values were defined as clinically implausible data (eg, SBP <80 or >220 mm Hg) or statistical outliers beyond the mean (SD 3). These values were first cross-checked against original records, such as monitoring device logs and paper questionnaires. Confirmed measurement errors were corrected, while unverifiable values were marked as missing.

Missing data were subsequently handled based on variable criticality. Participants missing ≥20% of critical measurements were excluded according to predefined criteria. For noncritical variables and partially missing critical variables (eg, single-time RWT), multiple imputation (MI) was used to generate 5 complete datasets, and the results were pooled for analysis.

To guarantee standardization and accuracy, the processes of data cleaning, merging, and handling missing values were independently repeated twice and cross-compared. Additionally, research variables were classified and assigned values based on high-quality literature. Furthermore, a complete-case analysis (CCA), strictly excluding participants with any missing covariate data, was performed as a sensitivity analysis to verify the robustness of the primary imputed findings.

Statistical Analyses

Modeling of Joint Trajectories of BP and HR

Analysis was performed using SAS (version 9.4; SAS Institute Inc) software, and GBMTM was completed through the SAS Proc Traj program, fitting joint trajectories of the 24-hour mean SBP, DBP, and HR. Models specifying 2-5 trajectory groups, with polynomial orders (constant, linear, and quadratic) for each variable within each group, were evaluated. The optimal number of groups and shape of trajectories were determined based on (1) minimization of Bayesian Information Criterion and Akaike Information Criterion, (2) average posterior probabilities of group membership >0.90 for each group, (3) entropy >0.80, (4) adequate group size (each group >5% of the sample), and (5) clinical interpretability.

Association Analysis of Joint Trajectories With ΔRWT

Echocardiography was performed at the first and third follow-up visits to obtain cardiac-related parameters of the patients. For missing values of relative ventricular wall thickness measured by echocardiography, 5 complete datasets were generated using MI, and the results were combined and analyzed. The relative change in RWT (ΔRWT) over the 6-month period between the first (3 months after baseline) and third (9 months after baseline) follow-up visits was defined as follows: ΔRWT was used as the dependent variable, and the data were distributed according to a non–normal distribution. The independent variables included 18 factors such as trajectory group, sociodemographic characteristics (age, gender, residency status, education level, and marital status), physical status (BMI), lifestyle (smoking and alcohol consumption), and treatment factors (medications used). Continuous variables were expressed as median (IQR), and comparisons between groups were made using the Kruskal-Wallis *H* test. Independent variables with *P*<.10 in

XSL•FO
RenderX

univariate analyses were simultaneously entered into the regression models using the Enter method. To ensure the robustness of our findings, 2 types of sensitivity analyses were conducted: first, by further adjusting for predefined biological confounders, specifically age and sex; and second, by performing a CCA, which strictly excluded participants with any missing covariate data to verify the consistency of the results. Multicollinearity was diagnosed using the variance inflation factor (VIF), which was considered to exist if the VIF exceeded 4. If multicollinearity was identified, variables with the highest VIF values were sequentially excluded (prioritizing those with weaker clinical relevance to the outcome), until all remaining variables had VIF ≤4, to ensure the stability of the multiple linear regression model. Furthermore, to justify the use of parametric linear regression for the non–normally distributed ΔRWT, residual diagnostics were systematically performed. The regression models were deemed statistically appropriate provided that the standardized residuals fell within the widely accepted threshold (approximately ±3), which confirms the absence of severe outliers and ensures that the assumptions of homoscedasticity and normality of residuals were adequately met. The statistical analyses were conducted using R (version 4.4.3; R Foundation for Statistical Computing) and SPSS (version 26.0; IBM Corp).

Sensitivity Analyses

To verify the robustness of the primary findings, sensitivity analyses were implemented. The first analysis forced established biological confounders (age and sex) into the regression models to address potential variable selection biases. The second used a CCA approach, strictly excluding any participants with missing covariate data to validate the stability of the imputed associations.

Antihypertensive Medications Assessment

Medication exposure during the 9-month follow-up was tracked via prescription refill records, categorizing the predominant regimen into drug classes including ARBs, CCBs, β-blockers, and combination therapies. Differences in medication utilization across the identified trajectory groups were evaluated primarily using chi-square tests.

Ethical Considerations

The study was approved by the Biomedical Research Ethics Review Committee of the School of Public Health, Sun Yat-sen University (approval number 2021-No.081). Written informed consent was obtained from all participants prior to enrollment. All study data were deidentified before analysis to ensure confidentiality. Participants received complimentary 24-hour ABPM and echocardiography assessments as compensation for their involvement in this research. No images or supplementary materials containing identifiable information are included in this manuscript.

Results

Joint Trajectories of 24-Hour Mean SBP, DBP, and HR

The relevant parameters during the fitting trajectory are shown in Table 1. Since the proportion of group members in the group number of 5 was less than 5%, the optimal number of groups for the trajectory model was chosen as 4. Then the number of longitudinal trajectory groups in patients with hypertension was determined to be 4, the order was adjusted, and the order of the final trajectory curve was determined by combining the morphology of the trajectory curves in the graphs and the *P* value, which were all determined to be 0th order (straight lines). All trajectory groups satisfied the following criteria [25]: absolute values of Bayesian Information Criterion and Akaike Information Criterion are minimum, proportion of group members >5%, entropy >0.80, and average posterior probability >0.90, indicating excellent model classification accuracy.

Table . Model selection criteria for group-based multitrajectory modeling of 24-hour mean blood pressure and heart rate. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025.

Number of groups	BIC[a]	AIC[b]	AvePP[c]	Proportion of group members (%)	Likelihood ratio	Entropy	Number of group members
2	−9443.33	−9406.35	96.59/97.48	52.47/47.53	−9384.35	0.85	110/103
3	−9348.48	−9294.70	93.00/93.53/95.88	27.67/48.99/23.37	−9262.70	0.87	59/105/49
4	−9280.14	−9209.56	96.32/92.45/92.31/94.29	27.90/28.18/25.56/18.36	−9167.56	0.88	59/60/54/40
5	−9247.55	−9160.15	95.30/91.01/89.27/92.14/91.13	26.71/29.73/25.30/14.27/3.99	−9108.15	0.89	57/64/53/31/8

[a]BIC: Bayesian Information Criterion.

[b]AIC: Akaike Information Criterion.

[c]AvePP: average posterior probability.

As described in Table 2, the 4 trajectory groups have the following characteristics: group 1: BP (123/68 mm Hg) and HR (67 bpm) were both at the lowest levels, but a diastolic pressure below 70 mm Hg may lead to inadequate organ perfusion (eg, myocardial ischemia and cerebral ischemia), particularly in older adult patients or those with concomitant coronary artery disease or carotid artery stenosis [45]; group 2: BP was well controlled (128/77 mm Hg), but the persistently elevated HR (80 bpm) suggested increased sympathetic nervous system activity, which may increase cardiovascular risk; group 3: BP was within the stage 1 hypertension range (134/82 mm Hg), with a normal HR (70 bpm); and group 4: BP was within the

XSL•FO
RenderX

stage 2 hypertension range (142/89 mm Hg), with high HR (84 bpm), indicating a decompensated state and the highest cardiovascular risk. The median pulse pressure across the 4 study groups ranged from 51 to 55 mm Hg. Although these group-level mean values were below the 60 mm Hg risk threshold specified by the European Society of Hypertension [45], this macro-level observation does not rule out the presence of elevated pulse pressure-related risk in individual patients.

Table . Baseline blood pressure, heart rate, pulse pressure, and main clinical characteristics of each joint trajectory group. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025.

Group	Group members, n	SBP[a], mm Hg	DBP[b], mm Hg	Pulse pressure, mm Hg	HR[c], bpm	Main characteristics
1	59	123	68	55	67	Lowest BP[d] and HR
2	60	128	77	51	80	Normal BP and high HR
3	54	134	82	52	70	Stage 1 hypertension range, normal HR
4	40	142	89	53	84	Stage 2 hypertension range, high HR

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HR: heart rate.

[d]BP: blood pressure.

The individual trajectories of BP and HR in 213 patients with hypertension (Figures 1A-1C) showed strong alignment with the joint trajectories identified by the GBMTM (Figures 2A-2C). Moreover, the empirical distributions of BP and HR at 9 months after baseline closely matched the model-predicted values for each trajectory group, confirming that the GBMTM reliably identified subgroups exhibiting distinct hemodynamic progression patterns. As illustrated in Figure 3, the 3D distribution of joint trajectories quantitatively distinguishes the spatiotemporal differentiation among the 4 trajectory groups within the multidimensional parameter space. This representation offers clinically valuable insights into the heterogeneous dynamic profiles across these physiological dimensions.

XSL•FO
RenderX

Figure 1. Individual longitudinal trajectories of 24-hour mean SBP, DBP, and HR. The spaghetti plots depict the individual hemodynamic progression of patients across the 4 identified trajectory groups (groups 1-4) over 6 months, with measurements taken at baseline (T0), 3 months (T1), 6 months (T2), and 9 months (T3). Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025. DBP: diastolic blood pressure; HR: heart rate; SBP: systolic blood pressure.

XSL•FO
RenderX

Figure 2. Joint trajectories of 24-hour mean SBP, DBP, and HR identified by group-based multitrajectory modeling (GBMTM). The line graphs illustrate the 4 distinct joint trajectory groups over 4 time points: baseline (T0), 3 months (**T1**), 6 months (T2), and 9 months (T3). Shaded bands around each trajectory represent the 95% CIs of the GBMTM-predicted values. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025. DBP: diastolic blood pressure; HR: heart rate; SBP: systolic blood pressure.

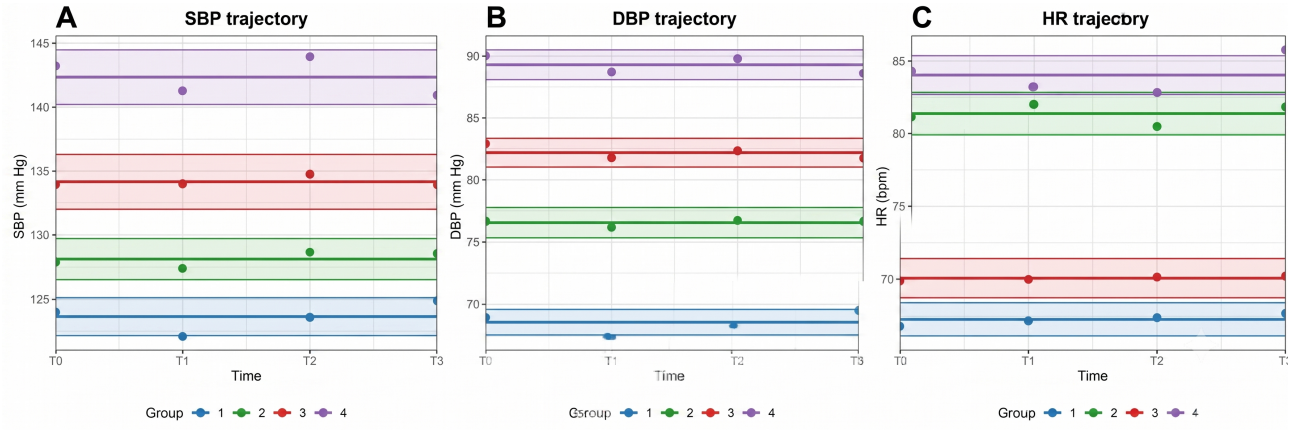
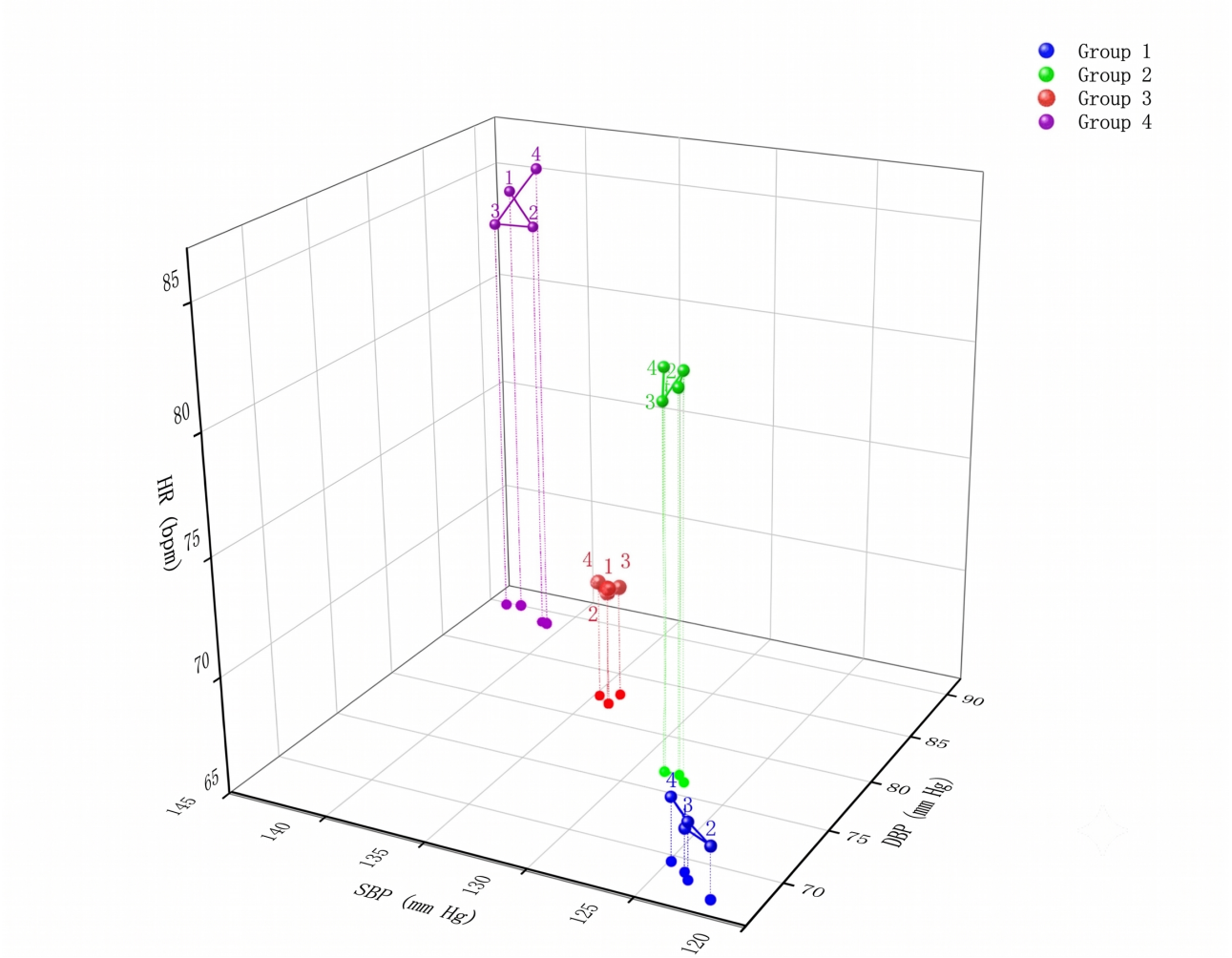


Figure 3. Three-dimensional (3D) spatiotemporal distribution of actual mean values for joint trajectories of SBP, DBP, and HR. The 3D scatter plots quantitatively distinguish the spatiotemporal differentiation among the 4 trajectory groups across 4 follow-up visits: baseline (1), 3 months (2), 6 months (3), and 9 months (4). Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025. DBP: diastolic blood pressure; HR: heart rate; SBP: systolic blood pressure.



Baseline Characteristics

The baseline characteristics of the study population are shown in Table 3. A total of 213 patients with hypertension were included in the study, 110 (51.64%) males and 103 (48.36%) females; their age ranged from 35 to 80 years, with a median age of 57 (IQR 52 - 65) years. The results of the Kruskal-Wallis

XSL•FO
RenderX

H test indicated that the 4 factors of trajectory group, wake-up time, dietary pattern, and history of diabetes mellitus met the prespecified threshold for inclusion in the multivariable model ($P<.10$).

XSL•FO
RenderX

Table . Baseline sociodemographic and clinical characteristics of the study population. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025[a].

Variable	Values, n	Relative change in RWT[b] (ΔRWT), median (IQR)	*H*	*P* value
Group			8.07	.05
1	59	0.02 (−0.03 to 0.07)		
2	60	−0.05 (−0.10 to 0.03)		
3	54	0.00 (−0.07 to 0.08)		
4	40	0.03 (−0.06 to 0.13)		
Sex			0.01	.92
Male	110	0.00 (−0.08 to 0.08)		
Female	103	0.00 (−0.06 to 0.06)		
Age (years)			2.12	.35
35-44	9	0.00 (−0.03 to 0.01)		
45-59	113	0.00 (−0.07 to 0.09)		
≥60	91	−0.02 (−0.08 to 0.05)		
Registered residence			0.08	.78
Local resident	57	−0.01 (−0.07 to 0.05)		
Non–local resident	156	0.00 (−0.07 to 0.08)		
Education level			0.59	.74
Primary school and below	79	0.00 (−0.07 to 0.05)		
Junior high school	106	0.00 (−0.07 to 0.08)		
High school and above	28	0.00 (−0.10 to 0.07)		
Employment status			0.08	.78
Full-time or part-time	108	0.00 (−0.07 to 0.08)		
Retired or unemployed	105	0.00 (−0.07 to 0.06)		
Marital status			0.10	.75
Married	207	0.00 (−0.07 to 0.08)		
Nonmarried	6	−0.01 (−0.06 to 0.04)		
Monthly household income (RMB)[c]			0.51	.78
<5000	126	0.00 (−0.07 to 0.07)		
5000-9999	61	0.00 (−0.07 to 0.07)		
≥10,000	26	0.00 (−0.06 to 0.08)		
BMI			1.86	.60
Underweight	2	−0.06 (−0.07 to −0.06)		
Normal	72	0.00 (−0.07 to 0.09)		
Overweight	95	−0.02 (−0.07 to 0.07)		
Obese	44	0.01 (−0.05 to 0.07)		
Smoking status			0.69	.71
Smoking	25	−0.02 (−0.07 to 0.08)		
Quit smoking	12	−0.03 (−0.09 to 0.03)		
Never smoked	176	0.00 (−0.07 to 0.07)		
Alcohol consumption			0.00	.99

XSL•FO
RenderX

Variable	Values, n	Relative change in RWT[b] (ΔRWT), median (IQR)	*H*	*P* value
Regular drinking (≥1/wk)	27	0.00 (−0.08 to 0.08)		
Less frequent drinking (<1/wk)	186	0.00 (−0.07 to 0.07)		
Wake-up time			3.04	.08
Before 7 AM	139	0.00 (−0.07 to 0.09)		
After 7 AM	74	−0.02 (−0.08 to 0.05)		
Bedtime			0.21	.65
Before 10 PM	48	0.00 (−0.08 to 0.12)		
After 10 PM	165	0.00 (−0.07 to 0.07)		
Sleep problems			0.01	.92
Yes	75	0.02 (−0.09 to 0.07)		
No	138	0.00 (−0.07 to 0.07)		
Dietary pattern			4.97	.08
Balanced diet	178	0.00 (−0.07 to 0.08)		
Meat-based diet	17	0.05 (−0.02 to 0.09)		
Plant-based diet	18	−0.05 (−0.13 to 0.03)		
History of diabetes			7.55	.02
Diabetes without medications	7	0.09 (−0.07 to 0.18)		
Diabetes with medications	33	−0.05 (−0.13 to 0.03)		
No diabetes	173	0.00 (−0.07 to 0.08)		
Antihypertensive medications			2.01	.57
ARB[d]	38	−0.03 (−0.09 to 0.05)		
CCB[e]	61	0.00 (−0.07 to 0.08)		
β-blocker	2	0.03 (0.03 to 0.03)		
Combination therapy	70	0.00 (−0.06 to 0.10)		
Sedentary behavior			1.01	.31
Yes	146	0.00 (−0.08 to 0.08)		
No	67	0.00 (−0.07 to 0.05)		

[a]The data for antihypertensive medication usage are based on available records (N=171). A total of 42 patients with unrecorded or missing medication data were excluded from this specific analysis.

[b]RWT: relative wall thickness.

[c]All household income values are presented in Chinese Yuan (RMB). The exchange rate was 1 RMB=0.138 US $ as of January 2025 (the study period).

[d]ARB: angiotensin receptor blocker.

[e]CCB: calcium channel blocker.

Multiple Linear Regression of ΔRWT

A multiple linear regression analysis was performed to model ΔRWT using the independent variables identified via univariate analysis. The corresponding results are presented in Table 4. Trajectory group, wake-up time, dietary pattern, and history of diabetes were converted to dummy variables, with group 4, wake-up time after 7 AM, balanced diet, and no diabetes set as reference categories. Multicollinearity diagnostics showed all VIFs <2.0.

XSL•FO
RenderX

Table . Multiple linear regression analysis of the independent associations of joint blood pressure and heart rate trajectories and baseline clinical covariates with the relative change in relative wall thickness (ΔRWT). Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025[a].

Variable	*b* [b]	*b*'[c]	SE	*t* test (*df*)	*P* value	95% CI of *b*
Group						
1	−0.016	−0.056	0.027	−0.598 (204)	.55	−0.070 to 0.037
2	−0.059	−0.205	0.027	−2.224 (204)	.03	−0.112 to −0.007
3	−0.008	−0.027	0.028	−0.292 (204)	.77	−0.063 to 0.047
4	Reference	N/A[d]	N/A	N/A	N/A	N/A
History of diabetes						
No diabetes	Reference	N/A	N/A	N/A	N/A	N/A
Diabetes without medications	0.042	0.060	0.049	0.845 (204)	.40	−0.056 to 0.139
Diabetes with medications	−0.067	−0.188	0.025	−2.681 (204)	.008	−0.116 to −0.018
Wake-up time						
After 7 AM	Reference	N/A	N/A	N/A	N/A	N/A
Before 7 AM	0.045	0.167	0.019	2.354 (204)	.02	0.007 to 0.083
Dietary pattern						
Balanced diet	Reference	N/A	N/A	N/A	N/A	N/A
Meat-based diet	0.054	0.116	0.032	1.663 (204)	.10	−0.010 to 0.117
Plant-based diet	−0.049	−0.107	0.032	−1.542 (204)	.13	−0.112 to 0.014

[a]F=3.392, P=.001, R^2=0.127, adjusted R^2=0.089. Model diagnostics: the assumption of normally distributed residuals was confirmed. The standardized residuals for this fully adjusted model ranged from −2.127 to 3.001, indicating the absence of severe outliers and justifying the use of parametric multiple linear regression for the non–normally distributed ΔRWT.

[b]*b*: unstandardized coefficient.

[c]*b*': standardized coefficient.

[d]N/A: not applicable.

Results indicated that trajectory group 2 (b=−0.059, 95% CI −0.112 to −0.007) and diabetes with medications (b=−0.067, 95% CI −0.116 to −0.018) were independently associated with a reduction in ΔRWT. Conversely, a wake-up time before 7 AM (b=0.045, 95% CI 0.007-0.083) was associated with an increase in ΔRWT (Table 4). The fully adjusted multiple linear regression model demonstrated a significant overall fit (F_{8204}=3.392; P=.001), explaining 12.7% of the variance in ΔRWT (R^2=0.127, adjusted R^2=0.089). This indicates that group 2, medication history, and wake-up time are independently associated with ΔRWT, although the overall variance explained by the model remains modest.

Sensitivity Analysis

To address potential variable selection and imputation biases, sensitivity analyses were performed by forcing age and sex into the regression models. In the sensitivity analysis using the fully adjusted model (Table 5), the association for group 2 remained statistically significant (P=.04). Furthermore, antidiabetic medication use (P=.009) and waking before 7 AM (P=.01) remained independently associated with ΔRWT. In this imputed model, age (P=.09) and sex (P=.66) did not reach statistical significance.

XSL•FO
RenderX

Table . Sensitivity analysis: fully adjusted multiple linear regression model for the association with change in relative wall thickness (ΔRWT). Sensitivity analysis forcing established biological confounders (age and sex) into the multiple linear regression model. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025[a].

Variable	*b*[b]	*b*'[c]	SE	*t* test (*df*)	*P* value	95% CI of *b*
Group						
1	−0.002	−0.008	0.029	−0.081 (202)	.94	−0.059 to 0.054
2	−0.056	−0.194	0.027	−2.082 (202)	.04	−0.109 to −0.003
3	−0.002	−0.008	0.028	−0.088 (202)	.93	−0.057 to 0.052
4	Reference	N/A[d]	N/A	N/A	N/A	N/A
History of diabetes						
No diabetes	Reference	N/A	N/A	N/A	N/A	N/A
Diabetes without medications	0.039	0.056	0.050	0.783 (202)	.44	−0.059 to 0.137
Diabetes with medications	−0.067	−0.187	0.025	−2.639 (202)	.009	−0.116 to −0.017
Wake-up time						
After 7 AM	Reference	N/A	N/A	N/A	N/A	N/A
Before 7 AM	0.051	0.188	0.019	2.606 (202)	.01	0.012 to 0.089
Dietary pattern						
Balanced diet	Reference	N/A	N/A	N/A	N/A	N/A
Meat-based diet	0.056	0.121	0.032	1.725 (202)	.09	−0.008 to 0.119
Plant-based diet	−0.047	−0.102	0.032	−1.484 (202)	.14	−0.110 to 0.016
Age (years)	−0.002	−0.129	0.001	−1.731 (202)	.09	−0.004 to 0.000
Sex						
Female	Reference	N/A	N/A	N/A	N/A	N/A
Male	−0.008	−0.032	0.019	−0.439 (202)	.66	−0.045 to 0.029

[a]F=3.045, P=.001, R^2= 0.141, adjusted R^2= 0.095. Model diagnostics: the assumption of normally distributed residuals was confirmed. The standardized residuals for this fully adjusted model ranged from −2.084 to 3.105, indicating the absence of severe outliers and justifying the use of parametric multiple linear regression for the non–normally distributed ΔRWT.

[b]*b*: unstandardized coefficient.

[c]*b'*: standardized coefficient.

[d]N/A: not applicable.

To further verify the robustness of these findings, a CCA was conducted (Table 6). In this model, the association for group 2 was attenuated to marginal significance (P=.06). This shift is primarily attributable to the reduction in statistical power when cases with missing data were excluded (n=196). Meanwhile, early waking (P=.004) and antidiabetic medication use (P=.01) robustly maintained their statistical associations. Furthermore, age emerged as an independent factor associated with ΔRWT (P=.047), whereas sex remained nonsignificant (P=.82).

XSL•FO
RenderX

Table . Sensitivity analysis: fully adjusted multiple linear regression model (complete case analysis) sensitivity analysis excluding missing data and forcing established biological confounders (age and sex) into the model. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025[a].

Variable	*b* [b]	*b*'[c]	SE	*t* test (*df*)	*P* value	95% CI of *b*
Group						
1	0.003	0.012	0.027	0.123 (185)	.90	−0.050 to 0.056
2	−0.048	−0.170	0.025	−1.908 (185)	.06	−0.097 to 0.002
3	−0.006	−0.022	0.026	−0.246 (185)	.81	−0.058 to 0.045
4	Reference	N/A[d]	N/A	N/A	N/A	N/A
History of diabetes						
No diabetes	Reference	N/A	N/A	N/A	N/A	N/A
Diabetes without medications	0.018	0.070	0.024	0.753 (185)	.45	−0.029 to 0.066
Diabetes with medications	−0.081	−0.231	0.032	−2.495 (185)	.01	−0.144 to −0.017
Wake-up time						
After 7 AM	Reference	N/A	N/A	N/A	N/A	N/A
Before 7 AM	0.053	0.199	0.018	2.897	.004	0.017 to 0.089
Dietary pattern						
Balanced diet	Reference	N/A	N/A	N/A	N/A	N/A
Meat-based diet	0.055	0.119	0.031	1.768 (185)	.08	−0.006 to 0.117
Plant-based diet	−0.045	−0.100	0.030	−1.506 (185)	.13	−0.104 to 0.014
Age (years)	−0.002	−0.143	0.001	−2.001 (185)	.047	−0.004 to −0.0003
Sex						
Female	Reference	N/A	N/A	N/A	N/A	N/A
Male	−0.004	−0.016	0.017	−0.228 (185)	.82	−0.038 to 0.030

[a]F=3.090, P=.001, R^2= 0.133, adjusted R^2=0.090. Model diagnostics: the assumption of normally distributed residuals was confirmed. The standardized residuals for this fully adjusted model ranged from –2.301 to 2.939, indicating the absence of severe outliers and justifying the use of parametric multiple linear regression for the non–normally distributed ΔRWT. This table presents the results of the complete-case analysis, excluding 17 participants with missing echocardiographic data (final analytical sample size, N=196).

[b]*b*: unstandardized coefficient.

[c]*b'*: standardized coefficient.

[d]N/A: not applicable.

Antihypertensive Medications Analysis

To investigate differences in the use of antihypertensive medications across trajectory groups, this study categorized patients into 5 groups based on their medication regimens over a 9-month period: ARBs, CCBs, β-blockers, combination therapy (≥2 medications), and unrecorded medications. Given the elevated HR in group 2 and the potential chronotropic effects of CCBs, we focused on comparing CCB usage rates across the 4 trajectory groups. Chi-square analysis revealed no statistically significant differences in overall antihypertensive drug distribution (χ^2_9=8.22; P=.51) or CCB usage rates (χ^2_3=3.30; P=.35) across groups (Table 7). Notably, group 4 demonstrated the highest CCB adoption rate (85.7%, predominantly amlodipine), exceeding group 2's rate (78.7%) by 7.0 percentage points, although this difference did not reach statistical significance. Amlodipine is primarily indicated for patients requiring BP reduction without sympathetic activation or metabolic abnormalities, and its HR-lowering effect is weaker than that of specific CCBs such as cilnidipine [46]. Furthermore, the VALUE trial demonstrated that amlodipine may induce reflex sympathetic activation via peripheral vasodilation, leading to a slight HR increase [47]. While this mechanism supports BP stability, it compromises effective HR control. The prevalent use of amlodipine in Zhongshan's 3 communities may contribute to sustained HR ≥80 bpm in these patients.

XSL•FO
RenderX

Table . Chi-square analysis of antihypertensive medication usage across the 4 joint trajectory groups over a 9-month period. Data were derived from a longitudinal cohort study conducted at 3 community health centers in Zhongshan City, Guangdong Province, China, among 213 patients with hypertension between November 2022 and January 2025[a].

Medications used	Group 1	Group 2	Group 3	Group 4	Chi-square (*df*)	*P* value
Antihypertensive medications	N/A[b]	N/A	N/A	N/A	8.22 (9)	.51
ARB[c]	12	11	11	4	N/A	N/A
CCB[d]	14	18	18	11	N/A	N/A
β-blocker	1	1	0	0	N/A	N/A
Combination therapy	18	18	14	20	N/A	N/A
CCB adoption	N/A	N/A	N/A	N/A	3.30 (3)	.35
Yes	31	37	33	30	N/A	N/A
No	14	10	11	5	N/A	N/A

[a]The data for antihypertensive medication usage are based on available records (N=171). A total of 42 patients with unrecorded or missing medication data were excluded from this specific analysis.

[b]N/A: not applicable.

[c]ARB: angiotensin receptor blocker.

[d]CCB: calcium channel blocker.

Discussion

Principal Findings

In this community-based cohort, GBMTM identified 4 distinct joint trajectory groups of BP and HR. Regarding the clinical implications, group 2 (characterized by normal BP and high HR) was significantly associated with a reduction in ΔRWT over 6 months, suggesting a potential improvement in LV geometry. BP in this group was well controlled and below the current guideline-recommended standard for hypertension of 130/80 mm Hg [45,48,49].

Furthermore, the use of antidiabetic medications in patients with hypertension with comorbid diabetes was significantly associated with a reduction in ΔRWT. This suggests that ameliorating insulin resistance may inhibit myocardial fibrosis and facilitate the repair of microvascular dysfunction [50-52], thereby providing evidence for the potential beneficial structural associations of these medications. Conversely, a wake-up time before 7 AM was significantly associated with an increase in ΔRWT, indicating that circadian dysregulation (manifested as early awakening) may elevate the risk of hypertensive cardiac remodeling by disrupting metabolic, endocrine, and BP regulation [42,53,54].

Comparison With Prior Work

In this study, unlike previous studies focusing solely on BP trajectories, we applied GBMTM to identify 4 distinct joint trajectories of 24-hour mean SBP, DBP, and HR in patients with hypertension. Previous studies have used the product of BP and HR (BP × HR), known as the double product, as an indicator of myocardial oxygen consumption. The double product serves as an indicator that simultaneously integrates BP and HR. It not only reflects the risk of angina pectoris episodes but also independently predicts all-cause mortality and cardiovascular mortality at rest, highlighting the necessity of simultaneously monitoring BP and HR in clinical assessment and risk prediction [55,56]. This metric increases with age in untreated populations and shows a U-shaped trend in treated populations. The dynamic changes of this index suggest that the synergistic elevation of BP and HR increases myocardial oxygen consumption, while their combined control may reduce related risks. This indicates that there is still significant room for improvement if the focus is solely on BP control, and combined management with other indicators (such as HR) may enhance control effectiveness [57].

Group 1 exhibited the lowest BP and HR levels; however, their DBP fell below 70 mm Hg. Consistent with the “J-curve” phenomenon in cardiovascular risk, clinical management for this phenotype should prioritize systolic BP control while closely monitoring DBP to avoid potential hypoperfusion [58].

Group 2 maintained optimal BP control (below the 130/80 mm Hg threshold) alongside improved ventricular structure, which aligns with findings that good BP control is associated with the potential reversal of LV hypertrophy [59]. However, their elevated resting HR (approximately 80 bpm) suggests persistent sympathetic activation. The structural improvement in ΔRWT in these patients may have preceded the reduction in HR, similar to the phenomenon of delayed autonomic modulation (HR variability) after structural improvement observed in patients with chronic thromboembolic pulmonary hypertension [60]. Despite this short-term structural improvement, long-term high HR may still pose potential risks to myocardial tissue, warranting caution. This study, through GBMTM, is the first attempt to reveal the cardiac remodeling characteristics of normal BP with high HR, providing a new perspective for risk stratification in patients with hypertension. These findings suggest that neglecting HR may lead to an underestimation of the potential beneficial associations with cardiac geometry in

XSL•FO
RenderX

some patients. Therefore, a combined assessment of BP and HR is necessary for evaluating risk in patients with hypertension.

Group 3 and group 4 represent escalating stages of pressure overload and autonomic decompensation, requiring aggressive, combined targeted therapies. Previous studies have shown that BP trajectories, by integrating long-term BP levels and the slope of changes, can predict cardiovascular events and mortality, providing important evidence for the risk assessment and intervention of CVDs. SBP trajectories have a stronger association with cardiovascular outcomes; furthermore, high-increasing SBP trajectories are also significantly correlated with elevated LVMI [61,62]. Previous studies have confirmed that elevated RWT is an independent predictor of CVD in patients with hypertension and diabetes [63]. This study reveals the synergistic association of dynamic interaction patterns between BP and HR with cardiac remodeling. Furthermore, regarding sleep patterns, the observed risk associated with early awakening aligns with chronobiological evidence that the morning BP surge and cortisol awakening response peak during the early morning window. Awakening prematurely triggers a shift toward sympathetic dominance and amplifies neuroendocrine stress, which are mechanistically linked to elevated vascular resistance. Finally, the observed potential beneficial associations of antidiabetic medications suggest that ameliorating insulin resistance may be linked to the inhibition of myocardial fibrosis and the improvement of microvascular dysfunction.

Our findings also highlight the clinical implication of 0th-order trajectories. The GBMTM algorithm identified 0th-order trajectories as the optimal fit for all 4 groups, indicating that the joint BP and HR profiles remained stable over the 9-month follow-up. This stability aligns with the clinical characteristics of the study cohort, which consisted of patients with chronic hypertension maintained on established medication regimens in a community setting, rather than newly diagnosed patients undergoing initial dose titration. Consequently, the 0th-order trajectories demonstrate that the identified hemodynamic phenotypes, such as the persistently elevated HR in group 2, represent chronic physiological states rather than transient measurements. Establishing this longitudinal stability provides the necessary methodological basis for assessing the cumulative longitudinal associations of these sustained profiles with continuous cardiac structural remodeling.

Strengths and Limitations

Our study possesses several key strengths. First, methodologically, this study applies GBMTM to a community-based hypertensive cohort to construct joint trajectories of SBP, DBP, and HR. This approach overcomes the limitations of traditional single-variable trajectory analyses, capturing the population heterogeneity of joint hemodynamic changes and identifying subgroups with distinct pathophysiological profiles. Second, by evaluating the progression of early cardiac remodeling across these joint trajectories, we explored the longitudinal determinants of structural changes. In contrast to previous studies that predominantly relied on cross-sectional associations, demonstrating the link between specific trajectory phenotypes and cardiac remodeling progression deepens our understanding of BP-HR interactions and provides longitudinal evidence for early intervention. Third, based on a Chinese community-dwelling cohort, this study highlights the optimal BP control range, the critical need for concurrent HR management, and the clinical significance of sleep interventions. These findings support the implementation of more precise, individualized hypertension management and cardiovascular risk reduction strategies within primary care settings.

However, several limitations of this study must be acknowledged. First, selection bias cannot be completely ruled out, as the sample was derived from a single community. Second, the specific roles of different classes of antidiabetic medications were not explored. Third, the follow-up period was relatively short. Although this duration was sufficient to observe early changes in ΔRWT, it is inadequate to fully evaluate long-term target organ damage and hard clinical end points. Fourth, echocardiographic measurements were obtained at the 3-month and 9-month follow-ups, rather than at baseline. Baseline imaging was omitted to reduce the logistical burden on primary care facilities during initial community recruitment. Consequently, the first 3 months functioned as a run-in period, during which approximately 200 participants were lost to follow-up. This attrition may have introduced selection bias, restricting the final cohort to patients with higher adherence. Furthermore, the shortened 6-month observation window (from month 3 to month 9) restricted the detectable magnitude of changes in ΔRWT. Because cardiac structural remodeling progresses slowly, this reduced time frame increases the likelihood that true biological changes are masked by the inherent measurement error of echocardiography, which may have attenuated the observed associations in our analysis. Nonetheless, the fact that significant correlations were identified despite these constraints suggests that the associations between high HR, BP trajectories, and cardiac remodeling are clinically noteworthy and warrant attention even within a limited observation window. Fifth, the utilization rate of β-blockers within our cohort was extremely low. Because β-blockers are the primary pharmacological agents for HR reduction, their underuse in this community setting likely contributed to the prevalence of the “normal BP, high HR” phenotype observed in group 2. Consequently, our findings primarily reflect the hemodynamic progression of patients with hypertension lacking specific HR-targeted interventions. This restricts the generalizability of our results to advanced clinical settings where β-blocker prescription rates are substantially higher. Sixth, regarding medication adjustments, we characterized each patient’s pharmacological exposure based on their most frequently prescribed regimen over the 9-month period. While this “predominant exposure” approach effectively captures the primary long-term maintenance therapy typical of community management, it inevitably simplifies the longitudinal pharmacological history. Transient dose titrations or brief medication switches were not continuously modeled as time-varying covariates in the GBMTM. Therefore, the acute hemodynamic impacts of short-term prescription fluctuations cannot be entirely ruled out. Seventh, the relatively modest sample size of our community cohort renders the multiple linear regression models somewhat sensitive to specific data-handling

XSL•FO
RenderX

techniques. This is evidenced by the minor discrepancies observed between the sensitivity analysis using imputed data and the strict CCA. Specifically, the association for group 2 attenuated from statistical significance to marginal significance in the CCA. This shift is primarily attributable to the reduction in statistical power when cases with missing data were excluded. Conversely, age emerged as an independently associated factor in the CCA but did not reach full statistical significance in the imputed model. This suggests that in a limited sample, the additional variance introduced by the MI process may dilute certain subtle associations, whereas complete-case exclusion may slightly alter the overall sample composition. These fluctuations highlight the necessity of interpreting the statistical associations cautiously and underscore the need for larger cohort studies to confirm these findings.

Future Directions

Future large-scale, multicenter prospective cohort studies with extended follow-up periods are needed to fully evaluate long-term cardiac target organ damage and hard clinical end points. Therefore, future research should extend the follow-up period beyond 2 years to better elucidate the temporal relationship between HR optimization and structural improvements, particularly for patients in group 2. Expanding recruitment across diverse community populations will help verify the external validity and generalizability of our findings. Additionally, using objective sleep measurement tools, such as polysomnography, would minimize self-reporting bias and improve the accuracy of sleep parameter assessments.

Conclusions

In this study, we applied GBMTM to longitudinal data on SBP, DBP, and HR in community-dwelling patients with hypertension, identifying 4 distinct joint trajectory groups of BP and HR. Notably, group 2 (characterized by normal BP and high HR) was significantly associated with a reduction in ΔRWT over 6 months, suggesting that achieving stable BP control is linked to a potential short-term attenuation of adverse LV geometric remodeling. However, it should be noted that this beneficial association attenuated to marginal significance in the CCA due to the reduction in statistical power. Additionally, concurrent antidiabetic medication use in patients with hypertension with comorbid diabetes was significantly associated with a reduction in ΔRWT, further supporting that these medications showed a beneficial association with the relative change in RWT. Conversely, waking up before 7 AM was associated with an increase in ΔRWT (Table 4), indicating that circadian rhythm disruption is involved in the progression of hypertensive cardiac remodeling.

These findings suggest the necessity to incorporate multidimensional BP and HR metrics to refine risk stratification and cardiac management in hypertension. Targeted interventions need to be explored in the future through long-term follow-up and the integration of multidimensional cardiovascular measurements.

Acknowledgments

The authors appreciate the efforts of all staff in data collection, data entry, and reporting in the monitoring communities (Qianlong community, Tanggan community, Guhe community, and Xinxu community). They acknowledge and thank the managers of the Health Monitoring Program in the above monitoring areas. Lun Li, Sanxiang Community Health Service Center of Zhongshan, Zhongshan 528463, China; Xibin Liu, Sanxiang Hospital of Zhongshan City, Zhongshan 528463, China; and Yunyue Zhao, Department of Cardiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510000, China, contributed to the revising of an early draft of the paper. During the preparation of this work, the authors used generative AI technologies (Gemini) strictly for the purpose of English language editing and polishing to improve the readability of the manuscript. After using this tool, the authors thoroughly reviewed and edited the content and take full responsibility for the final content of the publication.

Funding

This study was supported by grants from the Natural Science Foundation of Guangdong Province, China (numbers 2023A1515011951 and 2022A1515011237).

Data Availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to issues of participant confidentiality.

Authors' Contributions

JZ initiated the study and contributed to the methodological design and research coordination. QP, NZ, SY, JY, CD, YF, XZ, YZ, XH, JF, and HL collected, managed data, and conducted quality control. QP performed the statistical analysis and drafted the paper. QP, NZ, JF, and JZ provided an expert review and revision of the paper. JZ supervised the study. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol 2020 Apr;16(4):223-237. [doi: 10.1038/s41581-019-0244-2] [Medline: 32024986]
2. Mills KT, Bundy JD, Kelly TN, et al. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. Circulation 2016 Aug 9;134(6):441-450. [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
3. Global burden of disease 2021: findings from the GBD 2021 study. Institute for Health Metrics and Evaluation. URL: https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study [accessed 2025-07-16]
4. Global report on hypertension: the race against a silent killer. World Health Organization. URL: https://www.who.int/publications/i/item/9789240081062 [accessed 2025-07-16]
5. Pickering TG, Shimbo D, Haas D. Ambulatory blood-pressure monitoring. N Engl J Med 2006 Jun 1;354(22):2368-2374. [doi: 10.1056/NEJMra060433] [Medline: 16738273]
6. Kikuya M, Ohkubo T, Asayama K, et al. Ambulatory blood pressure and 10-year risk of cardiovascular and noncardiovascular mortality: the Ohasama study. Hypertension 2005 Feb;45(2):240-245. [doi: 10.1161/01.HYP.0000152079.04553.2c] [Medline: 15596571]
7. Dolan E, Stanton A, Thijs L, et al. Superiority of ambulatory over clinic blood pressure measurement in predicting mortality: the Dublin outcome study. Hypertension 2005 Jul;46(1):156-161. [doi: 10.1161/01.HYP.0000170138.56903.7a] [Medline: 15939805]
8. Gasowski J, Li Y, Kuznetsova T, et al. Is “usual” blood pressure a proxy for 24-h ambulatory blood pressure in predicting cardiovascular outcomes? Am J Hypertens 2008 Sep;21(9):994-1000. [doi: 10.1038/ajh.2008.231] [Medline: 18600212]
9. Hollenberg NK. Prognostic value of ambulatory blood pressure recordings in patients with treated hypertension. Curr Hypertens Rep 2004 Apr;6(2):80. [Medline: 15152650]
10. Rothwell PM, Howard SC, Dolan E, et al. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. Lancet 2010 Mar 13;375(9718):895-905. [doi: 10.1016/S0140-6736(10)60308-X] [Medline: 20226988]
11. Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. N Engl J Med 1990 May 31;322(22):1561-1566. [doi: 10.1056/NEJM199005313222203] [Medline: 2139921]
12. Koren MJ, Devereux RB, Casale PN, Savage DD, Laragh JH. Relation of left ventricular mass and geometry to morbidity and mortality in uncomplicated essential hypertension. Ann Intern Med 1991 Mar 1;114(5):345-352. [doi: 10.7326/0003-4819-114-5-345] [Medline: 1825164]
13. Gerdts E, Cramariuc D, de Simone G, Wachtell K, Dahlöf B, Devereux RB. Impact of left ventricular geometry on prognosis in hypertensive patients with left ventricular hypertrophy (the LIFE study). Eur J Echocardiogr 2008 Nov;9(6):809-815. [doi: 10.1093/ejechocard/jen155] [Medline: 18490279]
14. Ganau A, Devereux RB, Roman MJ, et al. Patterns of left ventricular hypertrophy and geometric remodeling in essential hypertension. J Am Coll Cardiol 1992 Jun;19(7):1550-1558. [doi: 10.1016/0735-1097(92)90617-v] [Medline: 1534335]
15. Oktay AA, Lavie CJ, Milani RV, et al. Current perspectives on left ventricular geometry in systemic hypertension. Prog Cardiovasc Dis 2016;59(3):235-246. [doi: 10.1016/j.pcad.2016.09.001] [Medline: 27614172]
16. Muiesan ML, Salvetti M, Monteduro C, et al. Left ventricular concentric geometry during treatment adversely affects cardiovascular prognosis in hypertensive patients. Hypertension 2004 Apr;43(4):731-738. [doi: 10.1161/01.HYP.0000121223.44837.de]
17. Devereux RB, Roman MJ, Paranicas M, et al. Impact of diabetes on cardiac structure and function: the strong heart study. Circulation 2000 May 16;101(19):2271-2276. [doi: 10.1161/01.cir.101.19.2271] [Medline: 10811594]
18. Eguchi K, Kario K, Hoshide S, Ishikawa J, Morinari M, Shimada K. Type 2 diabetes is associated with left ventricular concentric remodeling in hypertensive patients. Am J Hypertens 2005 Jan;18(1):23-29. [doi: 10.1016/j.amjhyper.2004.08.024] [Medline: 15691613]
19. McDonagh TA, Metra M, Adamo M, et al. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2023 Oct 1;44(37):3627-3639. [doi: 10.1093/eurheartj/ehad195] [Medline: 37622666]
20. Lundin M, Heiberg E, Nordlund D, et al. Prognostic utility and characterization of left ventricular hypertrophy using global thickness. Sci Rep 2023 Dec 20;13(1):22806. [doi: 10.1038/s41598-023-48173-7] [Medline: 38129418]
21. Fan JH, Wang JB, Wang SM, Abnet CC, Qiao YL, Taylor PR. Longitudinal change in blood pressure is associated with cardiovascular disease mortality in a Chinese cohort. Heart 2018 Nov;104(21):1764-1771. [doi: 10.1136/heartjnl-2017-312850] [Medline: 29691288]
22. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med 1998 Oct;55(10):651-656. [doi: 10.1136/oem.55.10.651] [Medline: 9930084]
23. Stokholm ZA, Bonde JP, Christensen KL, Hansen AM, Kolstad HA. Occupational noise exposure and the risk of hypertension. Epidemiology 2013 Jan;24(1):135-142. [doi: 10.1097/EDE.0b013e31826b7f76] [Medline: 23191997]

XSL•FO
RenderX

24. Nagin DS, Tremblay RE. Analyzing developmental trajectories of distinct but related behaviors: a group-based method. Psychol Methods 2001 Mar;6(1):18-34. [doi: 10.1037/1082-989x.6.1.18] [Medline: 11285809]
25. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010;6:109-138. [doi: 10.1146/annurev.clinpsy.121208.131413] [Medline: 20192788]
26. Li W, Jin C, Vaidya A, et al. Blood pressure trajectories and the risk of intracerebral hemorrhage and cerebral infarction: a prospective study. Hypertension 2017 Sep;70(3):508-514. [doi: 10.1161/HYPERTENSIONAHA.117.09479] [Medline: 28716992]
27. Yuan Z, Yang Y, Wang C, et al. Trajectories of long-term normal fasting plasma glucose and risk of coronary heart disease: a prospective cohort study. J Am Heart Assoc 2018 Feb 13;7(4):e007607. [doi: 10.1161/JAHA.117.007607] [Medline: 29440033]
28. Chu C, Liao YY, He MJ, et al. Blood pressure trajectories from childhood to youth and arterial stiffness in adulthood: a 30-year longitudinal follow-up study. Front Cardiovasc Med 2022;9:894426. [doi: 10.3389/fcvm.2022.894426] [Medline: 35845038]
29. Sabbah HN, Ilsar I, Zaretsky A, Rastogi S, Wang M, Gupta RC. Vagus nerve stimulation in experimental heart failure. Heart Fail Rev 2011 Mar;16(2):171-178. [doi: 10.1007/s10741-010-9209-z] [Medline: 21128115]
30. Heusch G. Heart rate in the pathophysiology of coronary blood flow and myocardial ischaemia: benefit from selective bradycardic agents. Br J Pharmacol 2008 Apr;153(8):1589-1601. [doi: 10.1038/sj.bjp.0707673] [Medline: 18223669]
31. Morcet JF, Safar M, Thomas F, Guize L, Benetos A. Associations between heart rate and other risk factors in a large French population. J Hypertens 1999 Dec;17(12 Pt 1):1671-1676. [doi: 10.1097/00004872-199917120-00003] [Medline: 10658932]
32. Palatini P. Role of elevated heart rate in the development of cardiovascular disease in hypertension. Hypertension 2011 Nov;58(5):745-750. [doi: 10.1161/HYPERTENSIONAHA.111.173104] [Medline: 21896939]
33. Zhang N, Pan Q, Yang S, et al. Utilizing circadian heart rate variability features and machine learning for estimating left ventricular ejection fraction levels in hypertensive patients: a composite multiscale entropy analysis. Biosensors (Basel) 2025 Jul 10;15(7):442. [doi: 10.3390/bios15070442] [Medline: 40710092]
34. Chinese Guidelines for the Prevention and Treatment of Hypertension Revision Committee. Chinese guidelines for the prevention and treatment of hypertension (2024 revised edition) [in Chinese]. Chin J Hypertens 2024;32(7):603-700. [doi: 10.16439/j.issn.1673-7245.2024.07.002]
35. Sagara K, Goto K, Maeda M, Murata F, Fukuda H. Medication adherence and associated factors in newly diagnosed hypertensive patients in Japan: the LIFE study. J Hypertens 2024 Apr 1;42(4):718-726. [doi: 10.1097/HJH.0000000000003661] [Medline: 38230627]
36. Wang J, Ma JJ, Liu J, Zeng DD, Song C, Cao Z. Prevalence and risk factors of comorbidities among hypertensive patients in China. Int J Med Sci 2017;14(3):201-212. [doi: 10.7150/ijms.16974] [Medline: 28367080]
37. Yip W, Wong TY, Jonas JB, et al. Prevalence, awareness, and control of hypertension among Asian Indians living in urban Singapore and rural India. J Hypertens 2013 Aug;31(8):1539-1546. [doi: 10.1097/HJH.0b013e328361d52b] [Medline: 23666423]
38. Cohn JN, Ferrari R, Sharpe N. Cardiac remodeling—concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. J Am Coll Cardiol 2000 Mar;35(3):569-582. [doi: 10.1016/S0735-1097(99)00630-0] [Medline: 10716457]
39. Muller JE, Stone PH, Turi ZG, et al. Circadian variation in the frequency of onset of acute myocardial infarction. N Engl J Med 1985 Nov 21;313(21):1315-1322. [doi: 10.1056/NEJM198511213132103] [Medline: 2865677]
40. Kario K, Pickering TG, Umeda Y, et al. Morning surge in blood pressure as a predictor of silent and clinical cerebrovascular disease in elderly hypertensives: a prospective study. Circulation 2003 Mar 18;107(10):1401-1406. [doi: 10.1161/01.cir.0000056521.67546.aa] [Medline: 12642361]
41. Somers VK, Dyken ME, Mark AL, Abboud FM. Sympathetic-nerve activity during sleep in normal subjects. N Engl J Med 1993 Feb 4;328(5):303-307. [doi: 10.1056/NEJM199302043280502] [Medline: 8419815]
42. Scheer F, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. Proc Natl Acad Sci U S A 2009 Mar 17;106(11):4453-4458. [doi: 10.1073/pnas.0808180106] [Medline: 19255424]
43. Li J. Introduction to the Self-Rating Scale of Sleep (SRSS) [in Chinese]. Chin J Health Psychol 2012;20(12):1851. [doi: 10.13342/j.cnki.cjhp.2012.12.040]
44. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of, Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2016 Apr;17(4):412-412. [doi: 10.1093/ehjci/jew041] [Medline: 26983884]
45. Mancia G, Kreutz R, Brunström M, et al. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension: Endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). J Hypertens 2023 Dec 1;41(12):1874-2071. [doi: 10.1097/HJH.0000000000003480] [Medline: 37345492]
46. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 1;39(33):3021-3104. [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]

XSL•FO
RenderX

47. Julius S, Kjeldsen SE, Weber M, et al. Outcomes in hypertensive patients at high cardiovascular risk treated with regimens based on valsartan or amlodipine: the VALUE randomised trial. Lancet 2004 Jun 19;363(9426):2022-2031. [doi: 10.1016/S0140-6736(04)16451-9] [Medline: 15207952]
48. Muntner P, Shimbo D, Carey RM, et al. Measurement of blood pressure in humans: a scientific statement from the American Heart Association. Hypertension 2019 May;73(5):e35-e66. [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]
49. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Hypertension 2018 Jun;71(6):1269-1324. [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]
50. Shah MS, Brownlee M. Molecular and cellular mechanisms of cardiovascular disorders in diabetes. Circ Res 2016 May 27;118(11):1808-1829. [doi: 10.1161/CIRCRESAHA.116.306923] [Medline: 27230643]
51. Frangogiannis NG. The extracellular matrix in ischemic and nonischemic heart failure. Circ Res 2019 Jun 21;125(1):117-146. [doi: 10.1161/CIRCRESAHA.119.311148] [Medline: 31219741]
52. Nesti L, Pugliese NR, Sciuto P, et al. Effect of empagliflozin on left ventricular contractility and peak oxygen uptake in subjects with type 2 diabetes without HEART disease: results of the EMPA-HEART trial. Cardiovasc Diabetol 2022 Sep 12;21(1):181. [doi: 10.1186/s12933-022-01618-1] [Medline: 36096863]
53. Zheng NS, Annis J, Master H, et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program. Nat Med 2024 Sep;30(9):2648-2656. [doi: 10.1038/s41591-024-03155-8] [Medline: 39030265]
54. Martino TA, Oudit GY, Herzenberg AM, et al. Circadian rhythm disorganization produces profound cardiovascular and renal disease in hamsters. Am J Physiol Regul Integr Comp Physiol 2008 May;294(5):R1675-R1683. [doi: 10.1152/ajpregu.00829.2007] [Medline: 18272659]
55. Robinson BF. Relation of heart rate and systolic blood pressure to the onset of pain in angina pectoris. Circulation 1967 Jun;35(6):1073-1083. [doi: 10.1161/01.cir.35.6.1073] [Medline: 6026199]
56. Inoue R, Ohkubo T, Kikuya M, et al. Predictive value for mortality of the double product at rest obtained by home blood pressure measurement: the Ohasama study. Am J Hypertens 2012 May;25(5):568-575. [doi: 10.1038/ajh.2012.3] [Medline: 22318510]
57. Asayama K, Hozawa A, Taguri M, et al. Blood pressure, heart rate, and double product in a pooled cohort: the Japan Arteriosclerosis Longitudinal Study. J Hypertens 2017 Sep;35(9):1808-1815. [doi: 10.1097/HJH.0000000000001399] [Medline: 28486272]
58. Vidal-Petiot E, Ford I, Greenlaw N, et al. Cardiovascular event rates and mortality according to achieved systolic and diastolic blood pressure in patients with stable coronary artery disease: an international cohort study. Lancet 2016 Oct 29;388(10056):2142-2152. [doi: 10.1016/S0140-6736(16)31326-5] [Medline: 27590221]
59. Ascher SB, de Lemos JA, Lee M, et al. Intensive blood pressure lowering in patients with malignant left ventricular hypertrophy. J Am Coll Cardiol 2022 Oct 18;80(16):1516-1525. [doi: 10.1016/j.jacc.2022.08.735] [Medline: 36229087]
60. Benza RL, Ghofrani HA, Grünig E, et al. Effect of riociguat on right ventricular function in patients with pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension. J Heart Lung Transplant 2021 Oct;40(10):1172-1180. [doi: 10.1016/j.healun.2021.06.020] [Medline: 34353714]
61. Allen NB, Khan SS. Blood pressure trajectories across the life course. Am J Hypertens 2021 Apr 2;34(3):234-241. [doi: 10.1093/ajh/hpab009] [Medline: 33821941]
62. Hao G, Wang X, Treiber FA, Harshfield G, Kapuku G, Su S. Blood pressure trajectories from childhood to young adulthood associated with cardiovascular risk: results from the 23-year longitudinal Georgia stress and heart study. Hypertension 2017 Mar;69(3):435-442. [doi: 10.1161/HYPERTENSIONAHA.116.08312] [Medline: 28093467]
63. Eguchi K, Ishikawa J, Hoshide S, et al. Differential impact of left ventricular mass and relative wall thickness on cardiovascular prognosis in diabetic and nondiabetic hypertensive subjects. Am Heart J 2007 Jul;154(1):79. [doi: 10.1016/j.ahj.2007.04.021] [Medline: 17584557]

Abbreviations

ABPM: ambulatory blood pressure monitoring
ARB: angiotensin receptor blocker
BP: blood pressure
CCA: complete-case analysis
CCB: calcium channel blocker
CH: concentric hypertrophy
CVD: cardiovascular disease
DBP: diastolic blood pressure
GBMTM: group-based multitrajectory modeling
HR: heart rate

XSL•FO
RenderX

LV: left ventricular
LVIDd: left ventricular internal dimension at end-diastole
LVMI: ventricular mass index
MI: multiple imputation
RWT: relative wall thickness
SBP: systolic blood pressure
VIF: variance inflation factor
ΔRWT: changes in relative wall thickness

Edited by A Coristine; submitted 04.Nov.2025; peer-reviewed by L Azevedo; revised version received 22.Jun.2026; accepted 22.Jun.2026; published 09.Sep.2026.

Please cite as:
Pan Q, Zhang N, Yang S, Ding C, Yin J, Fei Y, Zou X, Zheng Y, Huang X, Feng J, Lin H, Zhang J
Group-Based Multitrajectory Modeling Analysis of Joint Trajectories of Blood Pressure and Heart Rate and Their Association With Relative Change in Relative Wall Thickness in Patients With Hypertension: Community-Based Longitudinal Cohort Study
JMIR Cardio 2026;10:e87091
URL: https://cardio.jmir.org/2026/1/e87091
doi:10.2196/87091

© Qi Pan, Nanxiang Zhang, Shuo Yang, Chonglong Ding, Jianan Yin, Yi Fei, Xinyan Zou, Yongjun Zheng, Xiang Huang, Jiezhen Feng, Hai Lin, Jinxin Zhang. Originally published in JMIR Cardio (https://cardio.jmir.org), 9.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Health Care Utilization Among Patients With Coronary Artery Disease Using mHealth Wearable Devices: Feasibility Observational Study

Indranil Bardhan[1], PhD; Gari Clifford[2], PhD; Mujeeb Basit[3], MD; DuWayne Willett[4], MD; Tianjian Guo[5], PhD; Herman Taylor[6], MD

[1]McCombs School of Business, The University of Texas at Austin, 2110 Speedway, B6500, Austin, TX, United States

[2]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States

[3]Cardiology, Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, United States

[4]Cardiology, University of Texas Southwestern Medical Center, Dallas, TX, United States

[5]Kelley School of Business, Indiana University, Bloomington, IN, United States

[6]Cardiovascular Research Institute, Morehouse College, Atlanta, GA, United States

Corresponding Author:
Indranil Bardhan, PhD
McCombs School of Business, The University of Texas at Austin, 2110 Speedway, B6500, Austin, TX, United States

Abstract

Background: Although prior research has explored patient engagement with health care providers through software applications such as patient portals, the impact of patient engagement with mobile health (mHealth) technologies on subsequent utilization of health care resources remains poorly understood. Specifically, it remains unclear whether physiological data collected from wearable devices can be linked to future health care encounters across different modes of care utilization.

Objective: This pilot study was aimed at evaluating the feasibility of an mHealth prototype for analyzing cardiovascular disease outcomes and health care utilization among patients using digital health interventions. Specifically, we examined the feasibility of integrating physiological data from wearable devices to examine associations with future care utilization, offering early evidence for the effectiveness of patient engagement with mHealth platforms.

Methods: Patients with a history of coronary artery disease were recruited from the cardiology clinic of a large academic medical center. Before data collection, the study coordinator trained 20 patients on the use of Garmin devices, of whom 19 were included in the analyses. Data collected from the use of wearable devices were analyzed together with electronic health record data. Patient health outcomes were tracked from June 2019 to May 2021, including approximately 1500 health care encounters that encompassed hospital and outpatient visits, telehealth visits, and other encounters. Beyond establishing the feasibility of the mHealth platform, this work provided preliminary evidence of associations between (1) mHealth device use and future health care utilization and (2) physiological data and the frequency of future health care encounters.

Results: This study demonstrated the feasibility of an mHealth prototype for collecting and analyzing data to measure care utilization among patients with cardiovascular disease. Participants' use of the wearable mHealth device was associated with a statistically significant lower frequency of subsequent hospital and outpatient visits but greater utilization of other health care services, such as procedures, infusions, and medication refills, during the study period. However, none of these associations were statistically significant when the pretreatment period was restricted to the COVID-19 era. Minimum heart rate was associated with a higher frequency of future hospital and outpatient utilization, whereas activity time was associated with a higher frequency of procedural and refill visits.

Conclusions: This pilot study provides preliminary evidence that mHealth device use is associated with a reduction in future inpatient and outpatient visits and an increase in the number of visits for other health care services. It also indicates that patients' physiological data, such as minimum heart rate and activity time, are associated with the frequency of future health care encounters. These findings should be interpreted as exploratory signals rather than statistically definitive conclusions. This study provides proof of concept that an integrated mHealth platform can generate physiologically meaningful measures associated with subsequent care utilization, supporting its potential as a scalable tool for patient engagement and care coordination.

(JMIR Cardio 2026;10:e81922) doi:10.2196/81922

KEYWORDS

wearable devices; mHealth; patient engagement; physiological data; health care utilization

XSL•FO
RenderX

Introduction

This research was aimed at evaluating the feasibility of a novel mobile health (mHealth) technology prototype for studying cardiovascular disease by supporting patients in embracing behaviors that may improve cardiovascular health through innovative digital health strategies. We developed a novel, scalable approach to monitor and improve health behaviors among patients with cardiovascular disease using a smartphone app. This approach was designed to be unobtrusive and require little to no user input, except when participants chose to engage with the app.

The overall goal of this research was to improve patient engagement and care coordination with health care providers by using mHealth devices for chronic disease management. Specifically, our objectives were to (1) establish the feasibility of an mHealth prototype using "in the wild" data sourced from participants' use of wearable devices and (2) evaluate the relationships between patient engagement, use of mHealth devices, and future health care utilization.

Previous studies on patient engagement and care coordination have typically explored the use of tethered patient portals to assess patient engagement with care providers and its association with future health outcomes and utilization of health care resources [1]. Recent advancements in mobile technologies and cloud computing suggest that the use of smartphone apps is associated with a statistically significant increase in the 6-minute walk test distance [2,3]. Specifically, an observational study, based on a large cohort of over 30,000 users, reported significant associations between cardiovascular disease status (including peripheral artery disease, atrial fibrillation, and pulmonary arterial hypertension) and the 6-minute walking distance [4]. A related study showed that actigraphy data passively collected from smartphones can help differentiate decompensation events from euvolemic states among veterans with heart failure, with minimal cost and no user input [5]. Further, a randomized controlled trial involving 1070 patients with diabetes in Asia showed that adoption of an mHealth app was associated with improvements in both short-term metrics (reductions in blood glucose and glycated hemoglobin levels) and longer-term metrics, such as hospital visits and medical expenses [6].

However, a notable gap remains in our understanding of how patients with coronary artery disease (CAD) use smartphone-based health apps to monitor their lifestyles and engage with care providers [7]. Specifically, the association between patients' use of app-based wearable devices and patterns of future health care encounters, including hospital, emergency department, and outpatient visits, remains unclear and has not been empirically investigated [8,9]. For example, how does the use of wearable devices impact their future utilization of health care services, including hospital and emergency department visits as well as outpatient visits? Prior research has also not provided conclusive evidence regarding the relationship between the use of mHealth devices or apps that track patients' physiological data and the incidence of future health care encounters [10].

Our study aimed to address this research gap and demonstrate the feasibility of a novel mHealth prototype for patient engagement and care coordination in a cohort of 20 patients with CAD who were followed closely during the study period. Specifically, we examined the association between patients' use of the prototype mHealth wearable device and their overall engagement with health care providers. The mHealth device records essential physiological data, including blood pressure, heart rate, stress levels, sleep metrics, and physical activity. This streaming data feed is then matched with patients' electronic health record (EHR) data through a cloud-based platform for near real-time analytics. We assessed the viability of the mHealth prototype as a platform for patient-provider engagement by evaluating the relationships between physiological data collected through wearable devices and health care utilization in inpatient and outpatient settings. Beyond establishing the feasibility of the mHealth platform, we aimed to provide preliminary evidence of associations between mHealth device use and patient engagement and between specific types of physiological data and the frequency of future health care encounters. These analyses provide early insights into the potential association between mHealth interventions and health care utilization.

Methods

Methodology

We collected demographic, physiological, and clinical data from patients diagnosed with CAD. We then used statistical techniques to examine the associations between patients' physiological data collected through their engagement with wearable devices and their future utilization of health care resources. Our statistical analyses followed a 2-step approach, as described below.

In the first step, we characterized patients' patterns of health care engagement using EHR data. These data encompassed the timing and frequency of patients' interactions with health care providers, including inpatient, outpatient, and telehealth encounters, as well as other visits such as for medication refills, procedures, and infusions. The initial study protocol also proposed collecting additional information, including measures derived from text messages, food choices, and monthly vital signs, as well as administering a survey at the end of the study. However, disruptions caused by the COVID-19 pandemic prevented these data from being collected as planned. These protocol deviations are described in the addendum to Multimedia Appendix 1.

In the second step, we collected patients' physiological data, based on their use of mHealth wearable devices, and matched these data to their health care utilization records across the 4 types of encounters. We then fitted a generalized linear mixed model (GLMM, Equation 1) with a Poisson distribution and log link to conduct a prerecruitment-postrecruitment comparison aimed at evaluating whether patient enrollment in the mHealth platform was associated with changes in future health care utilization. The GLMM extends the standard generalized linear model framework by including patient-specific random intercepts along with fixed effects, accounting for within-patient

XSL•FO
RenderX

correlation arising from repeated weekly observations over time. In this setting, the random intercept captures unobserved, time-invariant heterogeneity across patients, allowing the model to isolate the association between mHealth enrollment and health care utilization while accounting for patient-specific factors.

$$(1) Visit_{it} \sim Poisson(\mu_{it}), \log(\mu_{it}) = \beta_0 + \beta_1 * Enroll_{it} + u_i u_i \sim N(0, \sigma_{u2})$$

Next, we conducted additional regression analyses to examine how patients' physiological activity measures, collected from their use of mHealth devices, are associated with health care utilization. Specifically, we regressed the frequency of each of the 4 types of health care encounters on physiological measures derived from their use of wearable devices, while incorporating patient fixed effects to control for unobserved, time-invariant heterogeneity across patients. The empirical model is presented in Equation 2:

$$(2) Encounters_{it} = f(Steps_{it}, ActiveTime_{it}, MinHeartRate_{it}, MaxStress_{it}, Sleep_{it}) + \alpha_i$$

where the dependent variable represents the number of health care encounters of each type (ie, hospital visits, outpatient visits, telehealth visits, and other visits) for patient *i* in week *t*. All explanatory variables were measured as weekly averages. Specifically, *Steps* was measured in thousands of steps, *ActiveTime* in hours, *MinHeartRate* in beats per minute, *MaxStress* on the Garmin 1 - 100 scale, and *Sleep* in hours. The term α_i represents patient fixed effects, which account for time-invariant, unobserved patient-specific characteristics (such as demographic data) without these variables explicitly included in the model.

As the dependent variable is a count, we used a Poisson GLMM regression model with conditional fixed effects to estimate the model presented in Equation 2. In addition to estimating models for contemporaneous associations within the same week, we estimated the model in Equation 2 using temporal leads of 1 to 4 weeks. To account for potential false discoveries, we applied the Benjamini-Hochberg false discovery rate (FDR) correction jointly across the complete family of 100 tests: 4 health care utilization outcomes (hospital, outpatient, telehealth, and other visits) × 5 physiological measures (total steps, total active time, minimum heart rate, maximum stress, and total sleep time) × 5 temporal alignments (week *t* through week *t*+4).

In summary, the overall methodology involved three steps: (1) time trend analyses of hospital and outpatient utilization before and after patient enrollment, (2) comparisons of health care encounters before and after patient recruitment, and (3) regression analyses of patient health care outcomes based on panel data on patient characteristics and physiological measures collected from wearable sensors and EHRs. Given the small sample size, our analyses were exploratory and intended to (1) test the viability of the mHealth prototype and (2) evaluate associations between physiological data collected through prototype use and future health care utilization.

Recruitment

We recruited patients with a history of CAD from the cardiology clinic of a large academic medical center in Texas. Initial screening for eligibility was based on data within the EHR system at the medical center. The eligibility criteria were as follows: (1) a record of at least one CAD-related visit in the cardiology clinic, (2) age greater than 21 years, (3) access to an Android smartphone, and (4) willingness to provide informed verbal consent based on the study protocol (Multimedia Appendix 1). Thus, eligible patients had a documented history of CAD based on their medical records. However, other information such as CAD severity and other clinical comorbidities, beyond baseline patient vitals, was not collected.

Next, the study research coordinator contacted patients who met the eligibility criteria and consented to participate. This resulted in a cohort of 20 patients, including 11 male and 9 female participants. All participants received a secure message from the study coordinator through the Epic MyChart app explaining the study requirements. First, the participants were required to use the Garmin vívosmart 4 wearable device, which was mailed to them at no cost, at all times during the study period. Second, they were asked to install 2 software apps on their Android smartphones based on guidance provided by the study coordinator. Third, they were required to participate in 2 video calls with the study coordinator to review the study objectives and requirements and receive assistance (if necessary) with installing the apps. Lastly, participants were asked to participate in several follow-up phone calls to answer any questions and verify that the smartphone apps were working properly. The Garmin vívosmart 4 was the primary mHealth device deployed to collect and process physiological data associated with participants' activities. Participants were required to use the wearable device at all times during the study period to ensure that their physiological data could be continuously collected and monitored.

Once participant training was completed, data collection commenced immediately. As the study was aimed at examining the association between the use of wearable mHealth devices and patient engagement and care utilization, measures were adopted to protect the integrity, confidentiality, and privacy of patient data. Data collected through the Garmin devices were initially stored on a Health Insurance Portability and Accountability Act (HIPAA)–protected Amazon Web Services (AWS) server and then transmitted to a secure server at the Texas Advanced Computing Center, where they were processed and analyzed for this study.

mHealth Platform

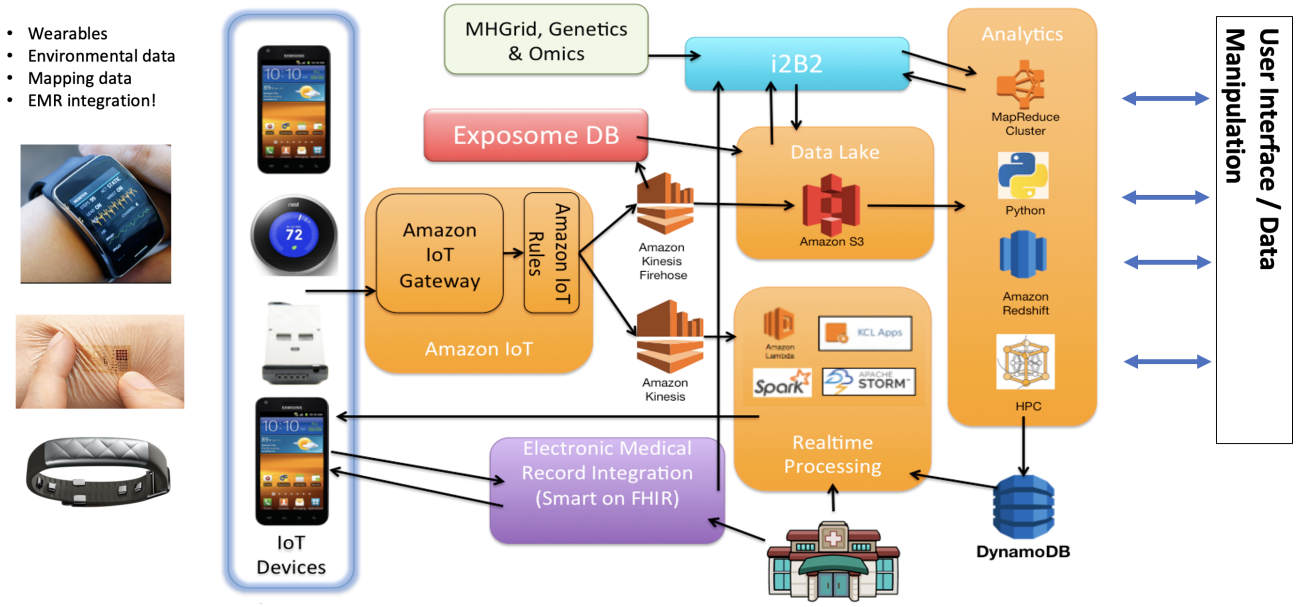
The mHealth platform enables long-term data collection from patients, including metrics related to functionality and fitness (physical activity patterns), autonomic function (heart rate variability), and sleep quality (latency and duration). The open-source app was built for smartphones using the Android operating system and has demonstrated high user compliance in prior research involving different clinical populations [11]. The prototype infrastructure leverages the HL7 Fast Healthcare Interoperability Resources (FHIR) standard, which provides representational state transfer (RESTful) access to clean, granular data. FHIR represents granular clinical concepts as a set of resources that can be accessed individually or in aggregate. The FHIR API supports queries at the individual patient and population level and enables the exchange of health records in highly readable XML and JSON formats. It also provides a large set of data models and vocabularies informed by real-world

XSL•FO
RenderX

implementation experience. We used FHIR to integrate the trace data generated from patients' use of the mHealth wearable devices with their EHR data, thus creating a longitudinal record of observational data for each patient enrolled in our study.

Figure 1 shows the overall architecture of the mHealth prototype platform developed and deployed to collect data from patients with CAD recruited at the academic medical center. Patients were monitored through web-connected or smartphone-connected devices, such as consumer fitness bands. The data were then piped to AWS nodes before being forwarded to the high-performance computing cluster at the research site and Amazon Simple Storage Service (S3). Researchers then analyzed the data using Amazon Elastic Compute Cloud (EC2). This infrastructure is highly scalable, always available, and extremely fault tolerant.

Figure 1. System design of the mobile health (mHealth) prototype platform developed to collect data from patients with coronary artery disease recruited into the study. The figure shows the technical components of the platform, as well as data flows among the patients' wearable devices, backend cloud computing services, electronic health record integration, and data processing using a high-performance computing (HPC) cluster at the research site. EMR: electronic medical record; FHIR: Fast Healthcare Interoperability Resources; IoT: Internet of Things.



Data Collection

We tracked patient health outcomes over a 2-year period, from June 1, 2019, to May 31, 2021. This period encompassed 18 months of data collected prior to the first wave of patient recruitment (which started in January 2021) and included over 5000 EHR encounters of patients' interactions with the health care system. After restricting these records to clinical encounters, approximately 1500 health encounters remained across the pretreatment and posttreatment periods combined (here, treatment refers to the use of the mHealth prototype wearable device after recruitment into the study). These encounters included inpatient hospitalizations, outpatient visits, telehealth visits, and other encounters such as visits related to medication refills, procedures, and infusions. Of these, approximately 200 health care encounters occurred during participants' use of the Garmin mHealth wearable devices.

The 18-month period between June 1, 2019, and December 31, 2020, served as the pretreatment period, allowing us to establish a baseline for evaluating patients' health care encounters and utilization patterns prior to their recruitment into the study. Although the data primarily documented the timing and type of each encounter, a smaller subset also included patient measurements, such as height, weight, blood pressure, and pulse rate. The posttreatment period began in January 2021, when the first patients were enrolled in the study. Because recruitment occurred on a rolling basis, individual start dates were staggered over the following months. Hence, our treatment cohort included patients recruited into the study who used the mHealth prototype wearable device, with each patient's posttreatment period defined relative to their enrollment date. Patient encounters with the health care system were documented in the EHR system of the cardiology clinic and matched to the wearable device data using a unique identifier.

Ethical Considerations

Morehouse College served as the overarching institutional review board (IRB) (Protocol #862899 - 14) for the participating academic institutions during the grant period. Patient recruitment and data collection were conducted at the University of Texas Southwestern Medical Center in accordance with the terms of the National Science Foundation grant (1636933) and prior IRB approval.

As patient recruitment and data collection were conducted primarily during and after the onset of the COVID-19 pandemic, the study received approval from the UT Southwestern Human Research Protection Program Reviewer (STU-2018 - 0355) to conduct all study activities virtually. Patients were recruited through messages sent via the Epic MyChart app, followed by virtual visits with the study coordinator. We obtained verbal informed consent from all participants, and verbal authorization pertinent to HIPAA was also requested. The study materials were updated to reflect these changes.

XSL•FO
RenderX

A copy of the informed consent document used as part of the study protocol is provided in Multimedia Appendix 1. Verbal informed consent was obtained by the study coordinator at the time of patient recruitment. An addendum describing deviations from the study protocol owing to disruptions caused by the COVID-19 pandemic is also included in Multimedia Appendix 1.

All study data were anonymized or deidentified for this research. No identifiable study participant data have been reported in this manuscript. Data were identified by unique study IDs rather than by any personal identifier. The code linking participant IDs to identifying information was stored separately from the study data and was accessible only to trained research staff directly involved in the project. Identifiers were destroyed 12 months after study completion. Electronic data were stored in secure, password-protected computer files. All research staff were trained in safeguarding confidential data and protecting human participants, as required at the study site.

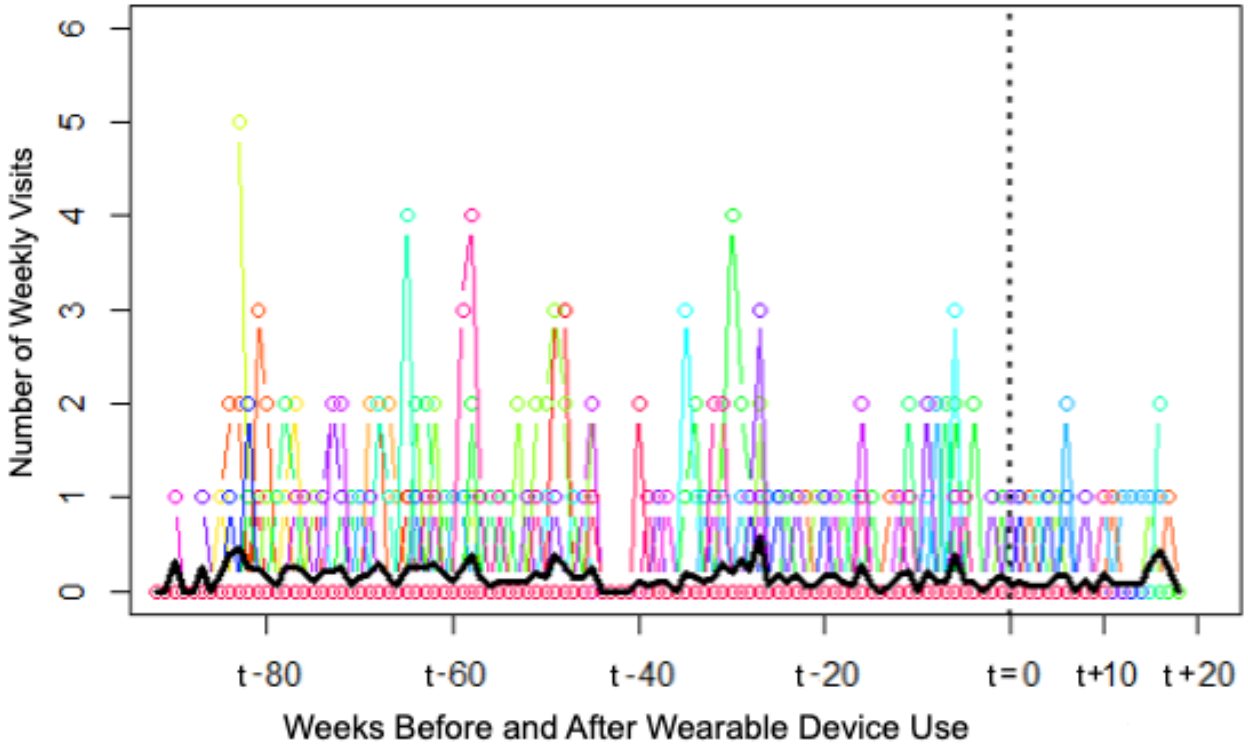
Participants did not receive monetary compensation for participating in the study. However, they were allowed to keep the Garmin wearable device provided at recruitment after the study ended.

Results

Statistical Analyses

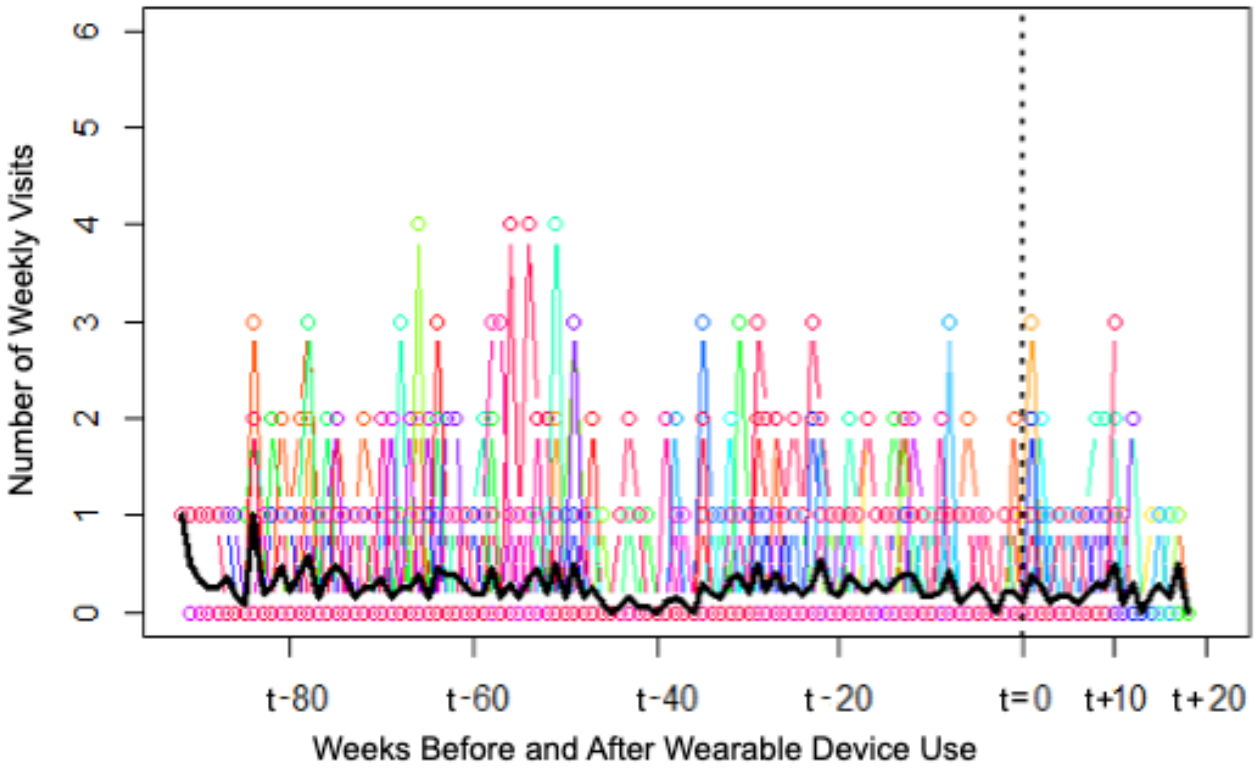
Across the 20 participants initially enrolled in the study, we assembled 1904 patient-weeks of EHR data, including 263 patient-weeks of postenrollment data and 176 patient-weeks of wearable device data. Because participants were enrolled on a rolling basis from January to May 2021, the duration of wearable device observations varied across participants, with a median of approximately 8 patient-weeks per participant and 25th and 75th percentiles of 7 and 11 patient-weeks per participant, respectively. This variation reflected the staggered schedule of enrollment, which led to shorter follow-up periods for participants enrolled later. During the initial data quality assessment, we excluded 1 participant whose data did not meet the study's reliability requirements. This exclusion resulted in a sample of 19 participants and 1803 patient-weeks of data, including 248 patient-weeks of postenrollment data. This sample was used for the descriptive time-trend analyses shown in Figures 2 and 3 and for subsequent pretreatment-posttreatment analyses.

Figure 2. Time trends of weekly hospital visits by patients. Weekly hospital visit counts for 19 patients with coronary artery disease enrolled in a longitudinal observational study of mobile health (mHealth) Garmin wearable device use. Time is measured in weeks relative to the start of device use, with *t*=0 denoting the week in which a patient first began wearing the device. Negative values indicate weeks prior to device use, and positive values indicate weeks following device use. Colored dotted lines represent weekly hospital visit counts for individual patients, and the solid black line represents average weekly hospital visit counts across patients. The vertical dotted line identifies the transition from the preparticipation period to the postparticipation period.



XSL•FO
RenderX

Figure 3. Time trends of weekly outpatient visits by patients. Weekly outpatient visit counts for the cohort of 19 patients with coronary artery disease enrolled in a longitudinal observational study of mobile health (mHealth) Garmin wearable device use. Visits are plotted relative to the start of wearable device use, with *t*=0 indicating the week of initial device adoption. Negative and positive time values correspond to preparticipation and postparticipation periods, respectively. Colored dotted lines denote weekly outpatient visit counts for individual patients, and the solid black line shows the average weekly outpatient visit counts across patients. The vertical dotted line identifies the transition from the preparticipation period to the postparticipation period.



To examine associations between physiological measures derived from mHealth device use and subsequent health care utilization, we matched the wearable data with EHR data by participant and patient-week. Patient-weeks with incomplete wearable data were excluded, reducing the 176 available patient-weeks to 131 complete, matched patient-weeks. Most exclusions resulted from missing sleep measures because participants removed their devices overnight despite being asked to wear them throughout the study period. These 131 patient-weeks constituted the sample used to summarize the weekly wearable measures in Table 1 and to estimate the regression models of physiological measures described below.

XSL•FO
RenderX

Table . Descriptive statistics of weekly patient physiological data[a].

	Mean (SD)	Median	Maximum	Minimum
Patient-week wearable device data				
Active kilocalories burned	1610.10 (977.76)	1698	4342	60
Steps (thousands)	36.78 (19.78)	37.52	72.61	3.38
Active time (hours)	14.58 (8.37)	14.59	34.18	1.30
Mean heart rate (bpm[b])	74.36 (5.67)	74.72	85.06	60.25
Median heart rate (bpm)	72.53 (5.80)	73	84	59
SD of heart rate (bpm)	11.19 (1.91)	11.34	16.14	7.03
Maximum heart rate (bpm)	133.56 (18.47)	133	193	102
Minimum heart rate (bpm)	51.33 (5.31)	52	62	38
Mean stress (Garmin scale, 1 - 100)	14.81 (4.28)	14.52	26.38	6.00
Median stress (Garmin scale, 1 - 100)	7.153 (5.57)	7	21	0
SD of stress	18.83 (3.74)	18.65	28.48	10.30
Maximum stress (Garmin scale, 1 - 100)	91.24 (5.80)	93	99	76
Minimum stress (Garmin scale, 1 - 100)	0 (0)	0	0	0
Total sleep time (hours)	41.19 (11.99)	45.45	61.35	3.98
Light sleep time (hours)	27.86 (8.10)	29.65	46.83	3.82
Deep sleep time (hours)	4.57 (3.48)	4.05	13.98	0
Pretreatment baseline (patient-level)				
BMI (kg/m²)	29.47 (6.34)	26.96	42.05	20.70
Systolic BP[c] (mmHg)	122.62 (9.90)	123.56	144.33	102.44
Diastolic BP (mmHg)	76.29 (6.81)	73.95	93.89	67.33
Heart rate (bpm)	74.64 (9.03)	73.33	91.14	57.86

[a]This table reports descriptive statistics for physiological measures collected from 19 patients with coronary artery disease during the posttreatment period, based on their use of a mobile health (mHealth) wearable device. The physiological measures include active calories burned, number of steps, active time, heart rate, stress, and sleep time. Posttreatment data were tracked and collected between January and May 2021 and included 131 patient-weeks of observational data recorded via the mHealth device. The table reports the mean, median, SD, maximum, and minimum values for each attribute. It also presents baseline vitals, including BMI, blood pressure, and heart rate, recorded at the time of patient enrollment.

[b]bpm: beats per minute.

[c]BP: blood pressure.

Table 1 presents descriptive statistics measured at 2 levels. The first panel summarizes data collected from weekly wearable device use using the sample of 131 matched patient-weeks. The second panel reports participants' baseline vital signs (body mass index, blood pressure, and heart rate) for the 19 participants. Although the original study protocol reported in Multimedia Appendix 1 specified monthly collection of vital signs, disruptions caused by the COVID-19 pandemic prevented these measurements from being collected regularly. These measures were too sparsely observed to be included as time-varying covariates. Therefore, each participant's baseline vitals at the time of enrollment were used for the analyses, as reported in the bottom panel of Table 1.

Care Utilization

Next, we evaluated time-variant patterns in the incidence of health encounters based on the number of weeks before and after patient participation. Figure 2 displays the average weekly hospital visit counts for the 19 participants. The horizontal axis shows time in weeks before and after study participation, with t=0 representing the week in which a patient first began using the wearable device. Because patient enrollment occurred on a rolling basis, t=0 was defined relative to each patient's individual start date rather than a shared calendar week, aligning all patients on a common study-time axis despite staggered enrollment. For example, t–20 and t+10 represent timelines 20 weeks before and 10 weeks after a patient's first use of the

XSL•FO
RenderX

wearable device, respectively. Similarly, Figure 3 shows the time plots of patients' weekly outpatient visits. Dotted lines in different colors represent weekly hospital visit counts for each patient, whereas the solid black line represents the average number of weekly hospital visits in Figure 2 and outpatient visits in Figure 3.

The top panel in Table 2 summarizes the mean weekly number of patient visits for each encounter type in the pretreatment and posttreatment periods. We compared differences in health care utilization across the four types of health care encounters, before and after patient enrollment, based on patients' use of the mHealth wearable device. The data presented in Table 2 pertains to 1803 patient-weeks, with 1555 and 248 patient-weeks corresponding to the pretreatment and posttreatment periods, respectively. This total is larger than the 131 patient-weeks reported in Table 1 as the analysis in Table 2 includes EHR data only (prior to matching these data with the mHealth wearable measurements).

Table . Pretreatment and posttreatment differences in patient health encounters[a].

Period/model	Hospital visits	Outpatient visits	Other visits	Telehealth visits	All visits
Mean encounters by period					
Pretreatment period	0.163	0.262	0.085	0.331	0.841
Posttreatment period	0.103	0.213	0.137	0.373	0.825
Overall (n=1803)	0.155	0.255	0.092	0.337	0.839
Pretreatment and posttreatment comparisons using GLMMs					
Poisson					
Coefficient	−0.540	−0.325	0.429	0.011	−0.124
z	−2.67	−2.16	2.28	0.10	−1.64
P value	.01	.03	.02	.92	.10
Poisson+patient baseline vitals					
Coefficient	−0.539	−0.323	0.441	0.013	−0.123
z	−2.66	−2.15	2.34	0.11	−1.64
P value	.01	.03	.02	.91	.10
Poisson, COVID-19 period					
Coefficient	−0.341	−0.139	0.330	−0.074	−0.083
z	−1.60	−0.87	1.66	−0.62	−1.04
P value	.11	.38	.10	.54	.30

[a]This table reports associations between adoption of the mHealth wearable device and weekly health care utilization for 19 patients with coronary artery disease. The analysis is based only on patients' electronic health record (EHR) data, including 1555 patient-weeks from the pretreatment period and 248 patient-weeks from the posttreatment period (for a total of 1803 patient-weeks). Visits are categorized as hospital, outpatient, telehealth, other (including refills, procedures, and infusions), or all visits combined. Each cell presents the posttreatment coefficient on the log-count scale from a generalized linear mixed model (GLMM) with a random patient intercept for repeated weekly observations; negative values indicate fewer posttreatment visits. Inference is based on 2-sided Wald *z* tests.

The bottom panel of Table 2 presents the estimated differences in weekly utilization for each encounter type between the pretreatment and posttreatment periods. We estimated GLMMs with a random patient intercept for repeated weekly observations. A negative coefficient indicates fewer visits in the posttreatment period, and its exponential, the rate ratio, shows the multiplicative change in expected weekly visits. The first row corresponds to the Poisson model, which regresses weekly visit counts on a posttreatment indicator and a random patient intercept. The random intercept accounts for the repeated weekly observations and unobserved, patient-specific heterogeneity. The second row (Poisson+patient baseline vitals) includes each patient's baseline vitals, including BMI, systolic and diastolic blood pressure, and heart rate as covariates. The Poisson GLMMs were estimated in R, version 4.3.0 (R Foundation for Statistical Computing), using the *glmmTMB* package, version 1.1.11.

Statistically significant associations were noted for hospital, outpatient, and other visits at the .05 significance level. The posttreatment coefficient for hospital visits was −0.540 (z=−2.67, P=.01), corresponding to a 42% reduction in expected weekly hospital visits relative to the pretreatment baseline. Outpatient visits declined by a comparable proportion, with a coefficient of −0.325 (z=−2.16, P=.03), corresponding to a 28% reduction. In contrast, other visits, including those for prescription refills, procedures, and infusions, increased, with a coefficient of 0.429 (z=2.28, P=.02), corresponding to a 54% increase. Differences in telehealth visits and total visits were not statistically significant.

As shown in the second row of the bottom panel of Table 2, inclusion of baseline patient vitals collected at the time of patient enrollment yielded quantitatively similar results across all outcomes. Overall, these findings suggest that use of the mHealth prototype wearable device was associated with a

XSL•FO
RenderX

reduction in inpatient and outpatient utilization, along with a concurrent increase in the utilization of other procedural health care services, including visits for preventative care.

Our estimates of changes in health care utilization were sensitive to the pretreatment, baseline period because it included several months of patient data collected before the onset of the COVID-19 pandemic. The COVID-19 pandemic represented an external shock, and its impact was difficult to disentangle from our pretreatment-posttreatment comparison analysis. This is because our pretreatment window (ie, June 2019-December 2020) spanned both the prepandemic and pandemic eras, whereas the posttreatment period (ie, January 2021 onward) fell almost entirely within the pandemic period. To address this concern, we conducted additional robustness analyses to reestimate the GLMMs while restricting the pretreatment period to the COVID-19 period only (ie, March 2020 onward). This ensured that both pretreatment and posttreatment analyses were based on a comparable pandemic-affected health care utilization environment.

The results for the COVID-19 period only are reported in the last row of Table 2. Statistical significance of the GLMM coefficients was attenuated for certain types of health care visits. For example, the significance of the estimated coefficients decreased from P=.01 to P=.11 for hospital visits, from P=.03 to P=.38 for outpatient visits, and from P=.02 to P=.10 for other visits. These results reflect the fact that the COVID-19 pandemic sharply suppressed hospital health care utilization before rebounding, which compressed the pretreatment-posttreatment differences when using only the pandemic-era data, thereby reducing statistical power rather than indicating the absence of an effect. Hence, we qualified the results reported in the bottom panel of Table 2 considering disruptions to care utilization patterns caused by the COVID-19 pandemic.

Physiological Data From Wearable Devices

Next, we analyzed wearable sensor data collected through participants' use of Garmin mHealth devices at the patient-week level. During the period for which data were available, coverage across patient-weeks was 95% for stress level and 75% for heart rate.

Five types of patient physiological data were considered in our analyses: number of steps taken, time spent physically active, heart rate, stress level, and sleep time. Sleep time included light sleep time, deep sleep time, remaining sleep time, and total sleep time.

Because we estimated conditional, fixed-effects Poisson models, the estimation sample for each lag specification was restricted to patients who contributed nonzero variation in the outcome within a particular lagged window. Patients with no within-patient variation in visit counts were excluded as they did not contribute any identifying information to the conditional likelihood. Consequently, the number of patient-weeks (N) and number of patients reported at the bottom of each regression table vary across columns. These values represent the effective estimation sample for the specific lag rather than the full study sample. The conditional, fixed-effects Poisson models were estimated in Stata/MP, version 17.0 (StataCorp), using the native *xtpoisson* command with patient-clustered robust standard errors.

Hospital Visits

Table 3 presents the Poisson regression results for hospital visits as the dependent variable. Benjamini-Hochberg FDR corrections were performed in Stata/MP, version 17.0 using native commands. The 100 2-sided Wald P values were ranked from smallest to largest, and the corresponding Benjamini-Hochberg–adjusted q values were calculated using the standard step-up procedure. Associations with $q<.05$ were considered statistically significant after FDR correction.

XSL•FO
RenderX

Table . Poisson regression results for hospital visits[a].

Independent variable	Week *t*	Week *t*+1	Week *t*+2	Week *t*+3	Week *t*+4
Steps (thousands)					
Coefficient	0.0233	−0.0708	0.0297	0.0269	−0.0839
z	0.33	−0.55	0.58	0.41	−0.91
P value	.74	.58	.56	.68	.36
Active time (hours)					
Coefficient	−0.1422	0.0227	−0.0087	−0.0963	0.2532
z	−0.73	0.07	−0.10	−0.43	0.89
P value	.47	.94	.92	.66	.37
Minimum heart rate (bpm[b])					
Coefficient	−0.0720	−0.0460	0.0200	*0.2610*	−0.0650
z	−1.63	−0.50	0.34	*3.94*	−1.13
P value	.10	.62	.73	*<.001*	.26
q	—[c]	—	—	*.004*	—
Maximum stress (Garmin scale)					
Coefficient	−0.0110	0.0910	0.0730	0.0390	0.0420
z	−0.20	1.23	0.88	0.29	0.48
P value	.84	.22	.38	.77	.64
Sleep time (hours)					
Coefficient	0.0404	−0.0351	−0.0429	0.0431	0.0985
z	0.94	−1.15	−1.02	0.95	2.36
P value	.35	.25	.31	.34	.02
N[d] (patient-weeks)	78	77	76	75	66
Patients[e]	8	8	8	8	7
Log-likelihood	−23.44	−21.43	−21.87	−18.59	−15.60

[a]This table summarizes results from conditional fixed-effects Poisson regression models to study the association between physiological activity measures and weekly hospital visits among patients with coronary artery disease. Hospital visit counts are measured weekly and regressed on contemporaneous and lagged physiological measures, including total steps, total active time, minimum heart rate, maximum stress, and total sleep time. Columns correspond to models using physiological measures from the same week *t* and 1-4 week leads (*t*+1 through *t*+4). All models include patient fixed effects to control for time-invariant patient characteristics and use robust standard errors clustered at the patient level. Coefficients represent log changes in the expected count of hospital visits. Statistical inference is based on Wald *z* tests from conditional fixed-effects Poisson models; exact 2-sided *P* values are reported in every cell. To account for potential false discoveries, Benjamini-Hochberg–adjusted q values are computed; associations surviving correction at q<.05 are shown in italics. Because patient fixed effects are eliminated through conditional likelihood, they are not estimated separately and do not consume residual degrees of freedom.

[b]bpm: beats per minute.

[c]Not applicable.

[d]N: conditional fixed-effects estimation sample (ie, patient-weeks contributing nonzero outcome variations within the lagged window).

[e]Patients: number of patients that exhibit within-patient identifying variation.

The regression results indicated a robust association between minimum heart rate and the number of hospital visits. Specifically, minimum heart rate in week *t* was positively associated with hospital visits in week *t*+3 (z=3.94, P<.001), and this association survived the Benjamini-Hochberg correction for multiple comparisons (q=.004), indicating it was unlikely to be a false discovery. Sleep time in week *t* was also positively associated with hospital visits in week *t*+4 at conventional significance levels (z=2.36, P=.02); however, this association did not survive correction for multiple testing and should therefore be interpreted with caution. Other variables, including number of steps taken, activity time, and maximum stress level, were not significantly associated with hospital visits at any time lead.

Although this study focused on the association between patient physiological data and care utilization, our results—particularly the robust finding for minimum heart rate—appear consistent with prior research reporting associations between higher heart rate and greater in-hospital mortality among patients [12,13],

XSL•FO
RenderX

even after accounting for differences in the severity of their health conditions.

Outpatient Visits

We estimated the model in Equation 2 using outpatient visits as the dependent variable. The Poisson regression results are presented in Table 4. Minimum heart rate in week *t* was positively associated with outpatient visits in week *t*+2 (*z*=3.36, *P*=.001), and this association survived the Benjamini-Hochberg correction for multiple comparisons (q=.026), indicating it was unlikely to be a false discovery. The positive coefficient suggests that elevated minimum heart rate was associated with higher outpatient utilization in the subsequent 2-week window. No other physiological measures, including steps taken, activity time, maximum stress, and sleep time, were significantly associated with outpatient visits robust to Benjamini-Hochberg correction, whether contemporaneously or at any of the 4 time lags examined.

Table . Poisson regression results for outpatient visits[a].

Independent variable	Week *t*	Week *t*+1	Week *t*+2	Week *t*+3	Week *t*+4
Steps (thousands)					
Coefficient	0.0940	−0.0316	0.1088	−0.0203	−0.1066
z	1.37	−0.40	1.23	−0.28	−1.23
P value	.17	.69	.22	.78	.22
Active time (hours)					
Coefficient	−0.4097	0.0801	−0.3484	0.1188	0.3207
z	−1.86	0.35	−1.68	0.56	1.45
P value	.06	.72	.09	.58	.15
Minimum heart rate (bpm[b])					
Coefficient	0.0960	−0.1270	*0.1370*	−0.0550	−0.0030
z	1.63	−1.52	*3.36*	−0.78	−0.04
P value	.10	.13	*.001*	.44	.96
q	—[c]	—	*.026*	—	—
Maximum stress (Garmin scale)					
Coefficient	0.0570	−0.0230	0.0400	−0.0250	−0.0970
z	0.87	−0.48	0.45	−0.43	−1.07
P value	.38	.63	.65	.67	.28
Sleep time (hours)					
Coefficient	0.0009	−0.0026	0.0134	−0.0307	0.0051
z	0.04	−0.18	0.36	−1.60	0.11
P value	.97	.85	.72	.11	.91
N[d] (patient-weeks)	85	91	87	92	89
Patients[e]	10	11	10	11	11
Log-likelihood	−39.92	−41.80	−37.27	−35.12	−33.23

[a]This table reports estimation results from conditional fixed-effects Poisson regression models examining the association between physiological activity measures and weekly outpatient visits among patients with coronary artery disease. Outpatient visit counts are measured weekly and regressed on contemporaneous and lagged physiological measures, including total steps taken, total active time, minimum heart rate, maximum stress, and total sleep time. Columns correspond to models using physiological measures from the same week *t* and from 1- to 4-week leads (*t*+1 through *t*+4). All models include patient fixed effects to control for time-invariant patient characteristics and use robust standard errors clustered at the patient level. Coefficients represent log changes in the expected count of outpatient visits. Statistical inference is based on Wald *z* tests from conditional fixed-effects Poisson models; exact 2-sided *P* values are reported in every cell. To account for potential false discoveries, Benjamini-Hochberg–adjusted q values are computed; associations surviving correction at q<.05 are shown in italics. Because patient fixed effects are eliminated through conditional likelihood, they are not estimated separately and do not consume residual degrees of freedom.

[b]bpm: beats per minute.

[c]Not applicable.

[d]N: conditional fixed-effects estimation sample (ie, patient-weeks contributing nonzero outcome variations within the lagged window).

[e]Patients: number of patients that exhibit within-patient identifying variation.

XSL•FO
RenderX

Our results regarding the association between minimum heart rate and frequency of hospital and outpatient visits are consistent with extant research. A recent study found that higher heart rate was associated with greater in-hospital mortality among patients in medical and surgery wards [12]. Among older patients hospitalized with heart failure, higher heart rate at discharge was associated with increased risks of death and rehospitalization, especially within the first 30 days after discharge [13]. Higher minimum heart rate may reflect underlying patient health risks, which could explain future hospital and outpatient visit utilization.

Telehealth Visits

We estimated the model using telehealth visits as the dependent variable. The Poisson regression results are presented in Table 5. Several coefficients were statistically significant at conventional levels: maximum stress in week t (z=–2.64, P=.008), minimum heart rate in week t+2 (z=2.47, P=.01), and active time in t+4 (z=2.34, P=.02). However, none survived the Benjamini-Hochberg FDR correction for multiple comparisons and were thus not interpreted as robust associations. Steps taken and sleep time were not significantly associated with telehealth visits. Given the lack of FDR-robust findings in this specification owing to the low statistical power of the sample, substantive conclusions could not be drawn regarding the relationship between physiological activity measures and telehealth utilization.

XSL•FO
RenderX

Table . Poisson regression results for telehealth visits[a].

Independent variable	Week *t*	Week *t*+1	Week *t*+2	Week *t*+3	Week *t*+4
Steps (thousands)					
Coefficient	−0.0168	0.0537	0.0728	0.0113	−0.0923
z	−0.34	1.04	1.35	0.21	−1.72
P value	.73	.30	.18	.83	.09
Active time (hours)					
Coefficient	−0.0825	−0.1805	−0.2007	0.1385	0.3709
z	−0.40	−1.16	−1.27	0.82	2.34
P value	.69	.25	.20	.41	.02
Minimum heart rate (bpm[b])					
Coefficient	0.0380	−0.1060	0.1610	0.0630	−0.1040
z	0.62	−1.44	2.47	0.85	−1.39
P value	.54	.15	.01	.40	.16
Maximum stress (Garmin scale)					
Coefficient	−0.1110	0.1020	0.0290	0.1520	−0.0070
z	−2.64	1.67	0.51	1.86	−0.13
P value	.008	.09	.61	.06	.90
Sleep time (hours)					
Coefficient	0.0682	−0.0434	0.0089	−0.0501	−0.0281
z	1.59	−1.53	0.36	−1.01	−1.02
P value	.11	.13	.72	.31	.31
N[c] (patient-weeks)	97	80	79	84	83
Patients[d]	11	10	10	10	10
Log-likelihood	−39.09	−36.51	−40.30	−39.16	−38.77

[a]This table summarizes results from conditional fixed-effects Poisson regression models examining the association between physiological activity measures and weekly telehealth visits among patients diagnosed with coronary artery disease. Telehealth visit counts are measured weekly and regressed on contemporaneous and lagged physiological measures, including total steps taken, total active time, minimum heart rate, maximum stress, and total sleep time. Columns correspond to models using physiological measures from the same week *t* and from 1 to 4 week leads, *t*+1 through *t*+4. All models include patient fixed effects to control for time-invariant patient characteristics and use robust standard errors clustered at the patient level. Coefficients represent log changes in the expected count of telehealth visits. Statistical inference is based on Wald *z* tests from conditional fixed-effects Poisson models; exact 2-sided *P* values are reported in every cell. To account for potential false discoveries, Benjamini-Hochberg–adjusted q values are computed; no associations survived FDR correction at q<.05. Because patient fixed effects are eliminated through conditional likelihood, they are not estimated separately and do not consume residual degrees of freedom.

[b]bpm: beats per minute.

[c]N: conditional fixed-effects estimation sample (ie, patient-weeks contributing nonzero outcome variations within the lagged window).

[d]Patients: number of patients that exhibit within-patient identifying variation.

Other Visits

Lastly, we estimated the model using other visits as the dependent variable. The Poisson regression results are presented in Table 6. These visits were procedural in nature and included medication refill requests and medical procedures, such as infusions. A strong, FDR-robust positive association was noted between active time and other visits 4 weeks later, in week *t*+4 (z=6.54, P<.001, q<.001), indicating that each additional hour of weekly active time was associated with a 34% increase in expected procedural and refill-related visits. This magnitude was likely inflated by the small identifying sample and should thus not be interpreted as a clinically meaningful effect. Several other coefficients exhibited conventional significance: minimum heart rate in week *t*+1 (P=.007), maximum stress in weeks *t*+1 and *t*+2 (P=.02 and P=.02, respectively), and sleep time in week *t*+1 (P=.006). However, none survived Benjamini-Hochberg correction and were thus not considered robust associations.

XSL•FO
RenderX

Table . Poisson regression results for other visits[a].

Independent variable	Week *t*	Week *t*+1	Week *t*+2	Week *t*+3	Week *t*+4
Steps (thousands)					
Coefficient	−0.0698	0.0376	0.0182	−0.0474	−0.0586
z	−0.33	0.35	0.38	−0.38	−1.37
P value	.74	.73	.70	.70	.17
Active time (hours)					
Coefficient	0.2495	0.0779	−0.1040	0.2795	*0.2957*
z	0.36	0.23	−0.41	0.74	*6.54*
P value	.72	.81	.68	.46	*<.001*
q	—[b]	—	—	—	*<.001*
Minimum heart rate (bpm[c])					
Coefficient	0.1720	−0.2160	−0.1140	−0.0330	0.0120
z	1.11	−2.71	−0.96	−0.36	0.21
P value	.27	.007	.34	.72	.83
Maximum stress (Garmin scale)					
Coefficient	−0.0330	0.2410	−0.1370	0.0530	0.0460
z	−0.37	2.42	−2.41	0.44	0.45
P value	.71	.02	.02	.66	.66
Sleep time (hours)					
Coefficient	0.1074	0.1672	0.0158	0.0125	−0.0771
z	0.88	2.74	0.36	0.28	−1.65
P value	.38	.006	.72	.78	.10
N[d] (patient-weeks)	58	57	66	55	64
Patients[e]	6	6	7	6	7
Log-likelihood	−14.03	−10.98	−18.27	−18.03	−19.56

[a]This table summarizes results from conditional fixed-effects Poisson regression models examining associations between physiological activity measures and weekly counts of other visits among patients with coronary artery disease. Other visits primarily include medication refill requests and medical procedures such as infusions. Other visit counts are measured weekly and regressed on contemporaneous and lagged physiological measures, including total steps, total active time, minimum heart rate, maximum stress, and total sleep time. Columns correspond to models using physiological measures from the same week *t* and from 1- to 4-week leads (*t*+1 through *t*+4). All models include patient fixed effects to control for time-invariant patient characteristics and use robust standard errors clustered at the patient level. Coefficients represent log changes in the expected count of other visits. Statistical inference is based on Wald *z* tests from conditional fixed-effects Poisson models; exact 2-sided *P* values are reported in every cell. To account for potential false discoveries, Benjamini-Hochberg–adjusted q values are computed; associations surviving correction at q<.05 are shown in *italics*. Because patient fixed effects are eliminated through conditional likelihood, they are not estimated separately and do not consume residual degrees of freedom.

[b]Not applicable.

[c]bpm: beats per minute.

[d]N: conditional fixed-effects estimation sample (ie, patient-weeks contributing nonzero outcome variations within the lagged window).

[e]Patients: number of patients that exhibit within-patient identifying variation.

Summary

The statistical analysis of health care utilization trends using explanatory regression models indicates that patients' physiological data are associated with their health care utilization, usually with time lags, although the strength of evidence varies across outcome types. The most robust findings across all four outcomes are the positive association between minimum heart rate and hospital visits after 3 weeks (Table 3) and the positive association between minimum heart rate and outpatient visits after 2 weeks (Table 4), both of which survive FDR correction. Similarly, we find a robust positive association between active time and other visits after 4 weeks (Table 6). In contrast, none of the associations identified for telehealth visits (Table 5) survive correction. Moreover, several associations across outpatient and other visit models that meet conventional significance criteria are not FDR-robust owing to the small sample size and should thus be interpreted with caution.

XSL•FO
RenderX

Overall, our research suggests that physiological data generated "in the wild" using mHealth wearable devices can help explain trends in future health care utilization, including the frequency of inpatient, outpatient, and other procedural encounters [12,14]. Nevertheless, the statistical inferences are based on a small sample, and the observed relationships should be treated as associational rather than causal.

Discussion

Principal Findings

Our study demonstrates the feasibility of a novel mHealth prototype that integrates physiological data collected through wearable devices with care utilization data collected from the EHR system of a large academic medical center. Specifically, this pilot study highlights the feasibility of a novel mHealth prototype in analyzing health care utilization by collecting and evaluating relevant measures of care utilization among patients with CAD.

Our research also demonstrates associations between patients' physiological data and future health outcomes, measured based on care utilization across different health care delivery settings. Specifically, this pilot study evaluates the relationship between mHealth use and patients' utilization of health care resources, such as inpatient, outpatient, telehealth, and other visits. The presented insights may support the triple aim of health care reform, namely, better individual and population health and lower costs, through the design of effective mHealth interventions to improve patient care coordination and health outcomes [15].

Our longitudinal observational study is based on a small panel of 19 patients with a history of CAD recruited from the cardiology clinic of a major academic medical center. The participants provided consent, met the study eligibility criteria, and fulfilled the study requirements. Data were collected through participants' use of an mHealth prototype wearable device, which recorded various types of physiological data, such as activity time, sleep behavior, stress, heart rate, and steps taken, over a period of 5 months. Our findings suggest that physiological data from mHealth wearable devices can be used to study associations with the frequency of future care utilization in outpatient and inpatient settings.

Within this small cohort of patients with CAD, the frequency of future hospital and outpatient visits during the study period was significantly lower than that during the prior 18-month period. Other visits, including those for medication refills, infusions, and medical procedures, were significantly higher during the study period. Although the relationship between patients' use of the wearable device and frequency of health encounters is associational, the simultaneous drop in hospitalizations and increase in other visits suggests an overall change in patients' awareness of their health conditions. The study participants are more likely to use wearable devices to monitor their health conditions which, in turn, is associated with a drop in the frequency of hospital visits but greater utilization of other types of health care services, such as procedures and prescription refills. The latter finding suggests greater utilization of preventative care, as medication refills and procedures underscore patients' propensity to adhere to their treatment regimen and medication schedules by undergoing regular tests and medical checkups.

Comparison With Prior Work

Our results provide preliminary evidence of significant associations between patient physiological measures and incidence of future health encounters [16]. Among the physiological measures collected from the mHealth wearable device, minimum heart rate and activity time emerge as the only explanatory variables with associations that survive FDR correction for multiple comparisons.

These associations point to a broader pattern of patient engagement with health care providers that spans multiple channels, including hospital and outpatient visits as well as procedural and refill-related encounters. Specifically, minimum heart rate is positively associated with subsequent hospital and outpatient visits, whereas activity time is positively associated with other visits (procedural and refill related). These findings suggest that different physiological signals may be informative for different types of care seeking behavior by patients.

Limitations

This study should be interpreted in the context of practical constraints common to early stage mHealth research. First, patient recruitment was limited by the clinical setting, geographic location, and technology requirements for wearable device deployment, resulting in a focused cohort drawn from a single cardiology clinic of an academic medical center.

Second, the number of patient-weeks with usable physiological data was limited by attrition and missing data, further emphasizing the exploratory scope of our analysis. Given the small sample size, this pilot study should be treated as an exploratory analysis of the association between wearable physiological signals and health care utilization.

Third, the study period overlapped with the COVID-19 pandemic, during which health care delivery and utilization patterns changed substantially. Increased reliance on telehealth, deferred inpatient care, and broader shifts in patient behavior during this period may have influenced the observed utilization trends alongside wearable device use and should be validated with additional data collected during the post-pandemic period. Notably, these conditions reflect the real-world environment in which mHealth wearable technologies are increasingly being implemented and evaluated.

Fourth, the study may be subject to selection bias as the users who agreed to participate may have been more engaged in managing their health than those who declined. A technology bias may also have been introduced as data were collected only from users of Android smartphones.

Lastly, disruptions caused by the COVID-19 pandemic resulted in deviations from the study protocol. Specifically, certain types of data, such as participants' food choices and text messages on the smartphone app, could not be tracked or collected, and some patients could not be observed for the entire 3-month duration, as initially planned. Inclusion of such data in future

XSL•FO
RenderX

studies would help improve the overall robustness of our preliminary findings.

Future Directions

Future studies should expand the scope of this study to increase the sample size and length of the study period to improve the overall robustness of our preliminary results. Future work can be aimed at examining the intensity and frequency of wearable device use as potential predictors of future health care utilization. Research can also be conducted on the determinants of mHealth use, as well as the potential role of patient demographic characteristics and social determinants of health in moderating the impact of physiological data on health care outcomes. Finally, identification of a similar group of patients who do not use mHealth wearable devices could help assess the relationship between patients' physiological data (collected through in-person encounters) and future care utilization patterns.

Conclusions

The pilot mHealth prototype for assessing patient engagement and care coordination is intended to provide a scalable platform and technological infrastructure that can support efforts to improve patient engagement with health care providers. Results derived using wearable device data collected "in the wild" from the mHealth prototype platform highlight the importance of patient physiological data in examining associations with future health outcomes based on the frequency of different types of health care encounters [16]. Specifically, we find that physiological data such as active time and minimum heart rate are associated with future hospital and outpatient visits. Our study underscores the importance of processing and monitoring trends within wearable device data and integrating these results to develop actionable alerts and clinical intelligence for remote patient monitoring [17].

Acknowledgments

IB gratefully acknowledges the financial support of the Foster Parker Centennial Professorship and the Charles and Elizabeth Prothro Regents Chair at the University of Texas at Austin.

HT and GC were funded by the National Science Foundation (NSF) award #1636933 "BD Spokes: Large-Scale Medical Informatics for Patient Care Coordination and Engagement" and the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1TR002378.

Generative AI Use

The authors confirm that no AI or generative AI tool was used to write any portion of this manuscript.

Funding

We gratefully acknowledge the National Science Foundation, grant 1636933 "Large-scale Medical Informatics for Patient Care Coordination and Engagement","Large-scale Medical Informatics for Patient Care Coordination and Engagement," for supporting the development and testing of the mobile health prototype and patient data collection. This work was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1TR002378.

Data Availability

Anonymized data were collected under institutional review board (IRB) approval from Morehouse College (IRB Protocol 862899 - 14) and the UT Southwestern Human Research Protection Program Office. The deidentified datasets analyzed in this study are not publicly available because of participant privacy considerations and IRB restrictions but are available from the corresponding author on reasonable request, subject to applicable institutional approval and data use requirements.

Authors' Contributions

The authors confirm that they have read and approved the manuscript, and they agree to be held accountable for the work performed therein. TG was responsible for the data analysis, GC was responsible for the mobile health prototype, HT was responsible for IRB approval, MB and DW were responsible for patient recruitment and data collection at the cardiology clinic, and IB and TG were responsible for interpretation of the results. All authors shared the tasks of reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol.
[DOCX File, 16 KB - cardio_v10i1e81922_app1.docx]

References

XSL•FO
RenderX

1. Bao C, Bardhan IR, Singh H, Meyer BA, Kirksey K. Patient-provider engagement and its impact on health outcomes: a longitudinal study of patient portal use. MIS Q 2020 Jun 1;44(2):699-724. [doi: 10.25300/MISQ/2020/14180]
2. Yu S, Chai Y, Chen H, Sherman SJ, Brown RA. Wearable sensor-based chronic condition severity assessment: an adversarial attention-based deep multisource multitask learning approach. MIS Q 2022 Sep 1;46(3):1355-1394. [doi: 10.25300/MISQ/2022/15763]
3. Leng L, Arévalo Avalos MR, Aguilera A, Lyles CR. The impact of an adaptive mHealth intervention on improving patient-provider health care communication: secondary analysis of the DIAMANTE trial. JMIR Mhealth Uhealth 2025 Jul 17;13:e64296. [doi: 10.2196/64296] [Medline: 40674692]
4. Kim DS, Schuetz N, Johnson A, et al. Unlocking insights: clinical associations from the largest 6-minute walk test collection via the my Heart Counts Cardiovascular Health Study, a fully digital smartphone platform. Prog Cardiovasc Dis 2025;89:45-52. [doi: 10.1016/j.pcad.2025.01.010] [Medline: 39884325]
5. Cakmak AS, Lanier HJ, Reinertsen E, et al. Abstract 15444: passive smartphone actigraphy data predicts heart failure decompensation. Circulation 2019;140(suppl 1). [doi: 10.1161/circ.140.suppl_1.15444]
6. Ghose A, Guo X, Li B, Dang Y. Empowering patients using smart mobile health platforms: evidence from a randomized field experiment. MIS Q 2022 Mar 1;46(1):151-192. [doi: 10.25300/MISQ/2022/16201]
7. Nambiar R, Bhardwaj R, Sethi A, Vargheese R. A look at challenges and opportunities of big data analytics in healthcare. Presented at: 2013 IEEE International Conference on Big Data; Oct 6-9, 2013; Silicon Valley, CA p. 17-22. [doi: 10.1109/BigData.2013.6691753]
8. Sohn A, Speier W, Lan E, et al. Assessment of heart failure patients' interest in mobile health apps for self-care: survey study. JMIR Cardio 2019 Oct 29;3(2):e14332. [doi: 10.2196/14332] [Medline: 31758788]
9. Bente BE, Wentzel J, Schepers C, et al. Implementation and user evaluation of an eHealth technology platform supporting patients with cardiovascular disease in managing their health after a cardiac event: mixed methods study. JMIR Cardio 2023 Mar 24;7:e43781. [doi: 10.2196/43781] [Medline: 36961491]
10. Yeung AWK, Kulnik ST, Parvanov ED, et al. Research on digital technology use in cardiology: bibliometric analysis. J Med Internet Res 2022 May 11;24(5):e36086. [doi: 10.2196/36086] [Medline: 35544307]
11. Osipov M, Behzadi Y, Kane JM, Petrides G, Clifford GD. Objective identification and analysis of physiological and behavioral signs of schizophrenia. J Ment Health 2015;24(5):276-282. [doi: 10.3109/09638237.2015.1019048] [Medline: 26193048]
12. Yamamoto M, Ohta Y, Sakuma M, Takeuchi J, Matsumoto C, Morimoto T. Association between heart rate on admission and in-hospital mortality among general inpatients: Insights from Japan Adverse Drug Events (JADE) study. Medicine (Baltimore) 2019 Apr;98(16):e15165. [doi: 10.1097/MD.0000000000015165] [Medline: 31008937]
13. Laskey WK, Alomari I, Cox M, et al. Heart rate at hospital discharge in patients with heart failure is associated with mortality and rehospitalization. J Am Heart Assoc 2015 Apr 22;4(4):e001626. [doi: 10.1161/JAHA.114.001626] [Medline: 25904590]
14. Bianchini E, Alborghetti M, Galli S, et al. Minimal clinically important difference of average daily steps measured through a consumer smartwatch in people with mild-to-moderate Parkinson disease: cross-sectional study. JMIR Mhealth Uhealth 2025 Jul 29;13:e64213. [doi: 10.2196/64213] [Medline: 40734369]
15. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041]
16. Mouttham A, Peyton L, Kuziemsky C. Leveraging performance analytics to improve integration of care. 2011 May 22 Presented at: ICSE11; Waikiki, Honolulu HI USA p. 56-62. [doi: 10.1145/1987993.1988005]
17. Reyna MA, Josef CS, Jeter R, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. Crit Care Med 2020 Feb;48(2):210-217. [doi: 10.1097/CCM.0000000000004145] [Medline: 31939789]

Abbreviations

AWS: Amazon Web Services
CAD: coronary artery disease
EC2: Elastic Compute Cloud
EHR: electronic health record
FDR: false discovery rate
FHIR: Fast Healthcare Interoperability Resources
GLMM: generalized linear mixed model
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
mHealth: mobile health
RESTful: representational state transfer
S3: Simple Storage Service

XSL•FO
RenderX

Edited by A Coristine; submitted 05.Aug.2025; peer-reviewed by CA Sanches, RSG Mahmoud; revised version received 30.Sep.2026; accepted 06.Oct.2026; published 13.Nov.2026.

Please cite as:
Bardhan I, Clifford G, Basit M, Willett D, Guo T, Taylor H
Health Care Utilization Among Patients With Coronary Artery Disease Using mHealth Wearable Devices: Feasibility Observational Study
JMIR Cardio 2026;10:e81922
URL: https://cardio.jmir.org/2026/1/e81922
doi:10.2196/81922

© Indranil Bardhan, Gari Clifford, Mujeeb Basit, DuWayne Willett, Tianjian Guo, Herman Taylor. Originally published in JMIR Cardio (https://cardio.jmir.org), 13.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Telehealth Intervention on Functional Status, Anxiety, Depression, and Rehospitalization Among Older Adults Undergoing Coronary Artery Bypass Grafting: Randomized Controlled Trial

Jirawan Mala[1], BNS, MSc; Usavadee Asdornwised[2], BSc, MSc, PhD; Kessiri Wongkongkam[2], BNS, MNS, PhD; Natkamol Chansatitporn[3], BSc, MSc, SCD; Punnarerk Thongcharoen[4], MD

[1]Doctor of Philosophy Program in Nursing Science (International Program), Faculty of Medicine Ramathibodi Hospital and Faculty of Nursing, Mahidol University, Bangkok, Thailand

[2]Department of Surgical Nursing, Faculty of Nursing, Mahidol University, 2 Wang Lang Road, Siriraj, Bangkoknoi, Bangkok, Thailand

[3]Faculty of Public Health, Mahidol University, Bangkok, Thailand

[4]Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Corresponding Author:
Usavadee Asdornwised, BSc, MSc, PhD
Department of Surgical Nursing, Faculty of Nursing, Mahidol University, 2 Wang Lang Road, Siriraj, Bangkoknoi, Bangkok, Thailand

Abstract

Background: Telehealth has shown promise in enhancing care transitions and physical health outcomes in patients with cardiovascular disease. However, limited studies have explored its effect on functional status, psychological health, and rehospitalization, specifically in older patients undergoing coronary artery bypass grafting (CABG).

Objective: This study aimed to evaluate the effectiveness of a telehealth intervention in improving functional status, reducing anxiety and depression, and decreasing rehospitalization rates compared with usual care among older patients undergoing CABG.

Methods: The study was a 2-arm parallel randomized controlled trial. This was conducted in 2 phases. Phase 1 was conducted in the cardiac surgical units at a university hospital in Bangkok, Thailand. Phase 2 involved following up with the participant at home 30 and 90 days after discharge. A total of 84 older adults undergoing CABG were randomly assigned to either the control group (n=42), which received usual care (discharge planning), or the intervention group (n=42), which received a telehealth intervention based on the transitional care model in addition to usual care. The telehealth intervention included home monitoring via the "Zip Heart" app and scheduled video consultations. The primary outcome was functional status, measured using the Thai version of the Enforced Social Dependency Scale. Secondary outcomes included anxiety and depression, assessed using the Thai Hospital Anxiety and Depression Scale, and rates of rehospitalization. Data were collected at baseline, 30, and 90 days after discharge. Analyses were conducted using an intention-to-treat approach, with missing outcome data handled using multiple imputation. Two-way repeated-measures ANOVA was used to evaluate group, time, and group-by-time interaction effects.

Results: A total of 84 participants were randomized and included in the intention-to-treat analysis (intervention group, n=42; control group, n=42). At baseline, there were no statistically significant differences between the two groups. Significant group-by-time interactions were observed for functional status scores ($F_{2,164}$=32.09, ηp^2=.28; P<.001), anxiety ($F_{2,164}$=20.22, ηp^2=.2; P<.001), and depression ($F_{2,164}$=16.81, ηp^2=.17; P<.001). The intervention group demonstrated significantly greater improvements in functional status and greater reductions in anxiety and depression at both 30 and 90 days after discharge compared to the control group (all P<.001). Additionally, rehospitalization rates were significantly lower in the intervention group at 30 days (Z=2.77; P=.006) and between 31 and 90 days post discharge (Z=2.31; P=.02).

Conclusions: The Telehealth intervention is effective in improving functional and psychological outcomes and reducing rehospitalization rates among older patients undergoing CABG. Integrating telehealth into usual care can support recovery and enhance continuity of care.

Trial Registration: Thai Clinical Trials Registry TCTR20230816008; https://www.thaiclinicaltrials.org/show/TCTR20230816008

(JMIR Cardio 2026;10:e81777) doi:10.2196/81777

KEYWORDS

telehealth; older adults; coronary artery bypass graft; transitional care model; functional status; anxiety; depression; rehospitalization

XSL•FO
RenderX

Introduction

Background

Coronary artery disease (CAD) is one of the leading causes of hospitalization and mortality worldwide, with older adults aged 60 years and older being the most affected [1]. Coronary artery bypass grafting (CABG) is the gold standard treatment for severe CAD and has been shown to improve survival, quality of life, and long-term cardiovascular outcomes by reducing recurrent myocardial infarction and the need for repeat revascularization [2,3]. This is particularly important given the increasing number of older patients requiring CABG, many of whom have multiple comorbidities, including hypertension, diabetes mellitus, and renal disease, which are common in older adults and require complex clinical management. Despite advancements in CABG, it remains highly stressful and traumatic for patients [4,5].

Older individuals undergoing CABG face many preoperative and postoperative physical and psychological challenges, which may adversely affect recovery and increase the risk of rehospitalization [4-8]. In Thailand, more than 6000 CABG procedures are performed annually [9], and most patients are older adults with preexisting comorbidities [10,11]. Rehospitalization within 28 days after discharge occurs in approximately 15.6% of patients and is commonly attributed to wound infections, fatigue, chest pain, and syncope [11].

Functional status is a key indicator of recovery after CABG, particularly in older patients. Older patients who underwent CABG have been reported to be particularly susceptible to decline in activities of daily living after surgery compared with their younger counterparts [12] and experience a slow recovery in functional status following surgery [6].

Psychological factors, particularly anxiety and depression, significantly influence recovery following CABG and are associated with delayed recovery, higher readmission rates, and increased mortality, especially in older adults [5,13]. These symptoms may persist or worsen after surgery due to the stress of surgery [4,5]. Preoperative anxiety and depression have been associated with higher mortality and poorer postoperative outcomes. Older patients undergoing CABG tend to experience higher levels of these symptoms than younger patients [5,14].

Inadequate postoperative care and ineffective communication may exacerbate psychological distress, resulting in poor care transitions and adverse outcomes [15,16]. System-level challenges, including workforce shortages, fragmented communication, and insufficient continuity of care, further contribute to care transition failures, increased adverse events, and higher rehospitalization rates [7,17-19]. Consequently, older adults with complex health conditions undergoing CABG are particularly vulnerable to poor postoperative outcomes.

Various programs, including standard discharge planning, cardiac rehabilitation, and patient education, are recommended to improve recovery and reduce rehospitalization after CABG [2]. Furthermore, nurses play a crucial role in guiding patients undergoing CABG and their family caregivers throughout the recovery process. However, usual care tends to be hospital-centered, with limited focus on home-based monitoring and postdischarge counseling.

Telehealth

Telehealth has emerged as an important strategy for postdischarge management after CABG, particularly during the COVID-19 pandemic. Telehealth interventions have been shown to enhance patient engagement and satisfaction, support symptom monitoring, and enable timely clinical responses, thereby promoting continuity of care and reducing rehospitalization [11,20-30]. These interventions, including video consultations, remote monitoring, and educational platforms, also contribute to improvements in physical and psychological recovery.

Telehealth interventions, particularly those incorporating mobile health apps, have shown promise in improving functional health outcomes in older patients with cardiac conditions through cardiac rehabilitation and home monitoring [31,32]. Furthermore, telehealth, based on the transitional care model (TCM) integrating with remote monitoring and scheduled video visits, has also improved medication adherence, self-care, and rehospitalization outcomes in patients with chronic cardiovascular conditions [33,34]. However, most telehealth studies have been conducted in developed countries using diverse technologies and methods [35,36]. In Asia, telehealth has been applied to rehabilitation for long COVID-19 and home-based exercise among patients with cardiometabolic multimorbidity [37-39]. Nevertheless, evidence specific to older patients undergoing CABG remains limited, and prior studies in Thailand have largely focused on rehabilitation and monitoring, with less emphasis on psychological outcomes and postdischarge consultation [11,40]. Barriers to telehealth adoption among older adults include technological limitations and physical impairments [41,42]. Conversely, individualized approaches, patient-centered education, and age-friendly design features can enhance user satisfaction and engagement [43]. Furthermore, family and social support are critical for successful implementation of telehealth [44]. This study incorporated family caregivers into the telehealth intervention to address these challenges, enabling them to assist older patients undergoing CABG with daily activities and care management during the postdischarge period.

Therefore, the study aimed to evaluate the effectiveness of telehealth intervention in improving functional status, anxiety, depression, and rehospitalization compared with usual care among older patients undergoing CABG.

Methods

Overview

This study used a 2-arm, parallel-group, randomized controlled trial with a pretest-posttest control group design and repeated measures. Participants were randomly assigned to either the intervention or control group using a 1:1 block randomization method. The care protocols for each group are outlined below.

XSL•FO
RenderX

Description of the Intervention

The intervention group received the telehealth intervention in addition to usual care. The telehealth intervention protocol was guided by the TCM [18] (Multimedia Appendix 1). The telehealth intervention was conducted by a researcher who was a specialist cardiac nurse with at least 5 years of experience in cardiac care. The care process started in the hospital and continued to the participants' homes based on the eight core TCM components: (1) screening, (2) engaging the older adults and caregiver, (3) managing symptoms, (4) educating and promoting self-management, (5) collaboration, (6) assuring continuity, (7) coordinating care, and (8) maintaining relationships. The telehealth intervention encompassed 2 phases of the care process.

Phase 1 involved discharge planning at the cardiac surgical units. During hospitalization, the researcher developed individualized care plans, provided health education through a booklet, and trained participants and caregivers in self-management and the use of the "Zip Heart" app. Vital signs (eg, heart rate and blood pressure) were measured using automated devices and manually entered into the app by participants or caregivers. Devices and internet access were provided as needed, and follow-up video calls were scheduled prior to discharge.

Phase 2 involved home monitoring via the "Zip Heart" app and video call visits, providing continuous postdischarge support and monitoring in the home setting. In the first month, the participants used the "Zip Heart" app to record their vital signs. They also received daily reminders and weekly video calls addressing symptoms, activities, and concerns, starting in the first week after discharge. Each call lasted 10 - 20 minutes, depending on the participant's recovery progress [25,45]. Video calls also enabled visual assessment of the participant's physical condition and environment. The researcher consulted with a multidisciplinary team to address specific needs. All relevant data were recorded in the integrated discharge planning screening and follow-up form. Furthermore, infographics on postoperative care and cardiac rehabilitation were shared weekly. In the second and third months, the participants received biweekly video calls and weekly infographics on secondary cardiac prevention and behavior modification. Although all participants were scheduled for at least 8 video call visits during the study, the researcher used clinical judgment to determine if additional visits were needed. The video call visits ended when the participant was able to resume daily activities and required no further support from the health care team. An interview guide developed for this study was used to minimize respondents' burden and avoid repetitive questioning. Participants in the intervention group continued to receive routine postoperative outpatient clinic follow-up as part of usual care. A representative telehealth intervention is presented in Figure 1.

Figure 1. Representative telehealth intervention components: (A) Zip Heart app interface; (B) Zip Heart mobile app installation; (C) home monitoring; and (D) video call visit.



Control Group

The control group received usual care by the cardiac nurse, including orientation to the surgical unit, health education, and standard discharge planning. Postoperatively, the cardiac nurse and a multidisciplinary team monitored recovery, managed pain, encouraged rehabilitation exercises, and prepared participants and caregivers for discharge. On the day of discharge, participants received standardized discharge instructions based on the D-METHOD framework (Diagnosis, Medications, Environment, Treatment, Health behaviors, Outpatient follow-up, and Diet). This framework included information on diagnosis, medications, home environment, treatment regimen, health behaviors, outpatient follow-up, dietary recommendations, wound care, and potential complications. Participants were advised to attend a follow-up appointment with their cardiac surgeon one week after discharge at the outpatient department (OPD) and were provided with the surgical unit's contact information for any health-related inquiries [11,46].

Sample Size Calculation

The sample size was calculated using a formula recommended for experimental studies with repeated measures [47], with a

XSL•FO
RenderX

significance level of .05 and a statistical power of 0.8. Functional status was selected as the primary outcome for this calculation. The estimation was informed by a previous study [11], which shared similar characteristics with this study and examined the effects of telehealth programs on functional status and rehospitalization rates among patients following CABG. In that study, functional status was measured using the Thai version of the Enforced Social Dependency Scale (Thai-ESDS) [48]. The SD of functional status scores in the population has been reported to range from 6 to 8.1. Four weeks after discharge, the experimental group had a mean functional status score of 13.59 (SD 3.35), compared to 17 (SD 5.98) in the control group, yielding a mean difference of 3.41. The corresponding effect size, calculated using Cohen d based on the group means and SDs, was 0.7, indicating a medium effect. Based on this effect size, the required sample size was estimated to be 33 participants per group. To account for anticipated attrition, the sample size was increased to 42 participants per group (84 in total), corresponding to an approximate attrition allowance of 21.4%.

Sampling

A probability sampling method was used in this study. Participants were randomly selected from patients aged 60 years and older who were admitted for their first elective isolated CABG and met the inclusion criteria. Eligible participants were identified through the hospital's surgical scheduling system, which served as the sampling frame. Patients were consecutively selected based on their scheduled surgery dates, according to the standard CABG surgical list, until the required sample size was reached.

Participant Recruitment and Enrollment

All patients admitted for first-time elective isolated CABG were approached by nurses in the Cardiac Surgical Units on the day of hospital admission. These patients were provided with detailed information about the study. Individuals who expressed interest in participating were subsequently assessed by the researcher to determine eligibility based on predefined inclusion and exclusion criteria. Eligible individuals were invited to participate and were provided with a participant information sheet. Sufficient time was given to address any questions before obtaining written informed consent.

Baseline Assessment

After providing written informed consent, all participants were assessed for baseline data by the researcher on the day of hospital admission. The assessments included (1) demographic and clinical characteristics, collected using the Health-Related Data Questionnaire (HRDQ); (2) functional status, measured using the Thai ESDS; and (3) levels of anxiety and depression, assessed using the Thai version of the Hospital Anxiety and Depression Scale (Thai HADS).

Randomization and Allocation

Participants were randomly assigned to the intervention or control group (n=42 per group) using 1:1 block randomization to ensure balanced group sizes. The randomization sequence was generated using a computerized random number generator with a permuted block size of four in Microsoft Excel, with codes "A" (control) and "B" (intervention). Allocation concealment was maintained using sequentially numbered, opaque, sealed envelopes prepared according to the randomization sequence. Envelopes were opened only after informed consent was obtained. Participant allocation details, including serial number and patient identification codes, were recorded by the investigator and stored securely.

Blinding Design

Blinding of treatment allocation was not feasible for either the nurses delivering the intervention or the participants due to the nature of the study design. Therefore, the trial was conducted using an open-label design. To minimize assessment bias, outcome data were collected by a research assistant who was blinded to group allocation, thereby ensuring observer blinding during data collection. However, a residual risk of unblinding remained, as participants may have inadvertently disclosed their group allocation during interviews.

Study Design

This study was a single-center RCT with a pretest-posttest control group design and repeated measures. The trial evaluated the effectiveness of the telehealth intervention in addition to usual care, compared to usual care only, among older adult patients. This study was conducted in accordance with CONSORT (Consolidated Standards of Reporting Trials; checklist provided in Checklist 1) guidelines for randomized trials involving 2 groups[49] .

Study Setting and Population

The study was conducted in 2 phases. Phase 1 took place in the cardiac surgical units of a university hospital in Bangkok, Thailand, between September 2023 and June 2024. Phase 2 involved home-based follow-up assessments at 30 and 90 days after discharge. The target population consisted of older adults undergoing CABG. All patients diagnosed with CAD who were scheduled for elective CABG and met the inclusion criteria were eligible and subsequently enrolled in the study.

Inclusion and Exclusion Criteria

The inclusion criteria were patients aged 60 years or older and scheduled for first-time elective isolated CABG, those classified as functional class I-III according to the New York Heart Association (NYHA) classification [50], those able to access and use a smartphone with an Internet connection, either independently or with assistance from their primary family caregiver, those able to communicate in Thai, those without cognitive impairment, defined as a score of ≥3 on the Thai version of the Mini-Cog [51], and those having a primary family caregiver willing to participate in the study. The exclusion criteria were patients with severe dementia, Alzheimer disease, delirium, terminal illness, or psychiatric disorders diagnosed by a physician. Given the potential for physical or sensory impairments, activity limitations, and increased fall risk among older patients undergoing CABG, their primary family caregivers were included in the intervention to assist with daily activities.

Participants did not withdraw from the study; instead, some participants discontinued the telehealth intervention based on predefined criteria, including admission to a nursing home,

XSL•FO
RenderX

refusal to continue participation after enrollment, failure to report physical measures or symptoms for five consecutive days with unsuccessful telephone contact for 3 consecutive days, intraoperative cardiac arrest, or severe postoperative complications such as stroke or renal failure requiring dialysis. All participants were retained in the final analysis.

Treatment Fidelity

The telehealth intervention protocol was validated by a panel of cardiovascular experts to ensure treatment fidelity. The intervention was delivered by the researcher, a specialist cardiac nurse with over 5 years of experience in cardiovascular care, in collaboration with a multidisciplinary team. To maintain adherence to the intervention protocol and ensure quality control, the researcher received weekly supervision from her academic advisor throughout the study. A research assistant was a cardiac nurse who was independent of the intervention. She was trained to collect data at 30 and 90 days after discharge, and she was supervised during the data collection process by the researcher to prevent bias. Additionally, a pilot study was conducted prior to the main data collection to assess the feasibility and acceptability of the intervention; the pilot data were not published and were not included in the main analysis.

Measurement

Four instruments were used for data collection: the HRDQ, the Thai ESDS, the Thai HADS, and the Number of Rehospitalizations Questionnaire. All instruments were validated by three clinical experts using the Content Validity Index (CVI) based on expert ratings of item relevance. Furthermore, Cronbach α coefficient was calculated using IBM SPSS Statistics version 18 to assess the instruments' reliability. The participants were provided with standardized instructions on how to complete the questionnaires.

HRDQ

The researcher developed the HRDQ to collect demographic characteristics, including age, gender, marital status, and education, and health-related information at baseline, such as CABG type, operative time, NYHA classification, and hospital length of stay (LOS). The CVI of the HRDQ was 1. Internal consistency reliability was not assessed because the items are descriptive and non-scaled.

Thai ESDS

The ESDS was originally developed by Benoliel [52], refined for use in patients with chronic disease by McCorkle et al [53], and translated into Thai by Asdornwised [48]. The Thai ESDS measures changes in the functional status of patients with various illnesses, including cancer, CAD, and stroke, by evaluating performance before and after illness. The Thai ESDS consists of 10 items covering two domains: personal and social competence. The total social competence score ranges from 4 to 15, whereas the total functional status score ranges from 10 to 51, with higher scores indicating greater enforced dependency. Functional status scores are categorized as follows: 10 - 23 indicates high independence, 24 - 37 indicates moderate independence, and 38 - 51 indicates low independence. The reliability score of the Thai ESDS was 0.71. The internal reliability (Cronbach α) was 0.80.

Thai Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) was originally developed by Zigmond and Snaith [54] and was translated into Thai by Nilchaikovit [55]. The HADS assesses both anxiety and depression levels. The HADS consists of 14 items divided equally into two subscales: 7 items for anxiety and 7 items for depression. Each item is rated on a 7-point scale ranging from 0 to 3, with a total score of 0 - 21 for each subscale. Scores on the Hospital Anxiety and Depression Scale–Anxiety subscale (HADS-A) are categorized as follows: 0 - 7 indicates no anxiety, 8 - 10 indicates mild anxiety, and 11 - 14 indicates moderate anxiety. Similarly, scores on the Hospital Anxiety and Depression Scale–Depression subscale (HADS-D) are categorized as follows: 0 - 7 indicates no depression, 8 - 10 indicates mild depression, and 11 - 14 indicates moderate depression. The Cronbach α coefficient for the HADS in Thai patients with CAD was 0.83 [40], whereas the Cronbach α for the anxiety and depression subscales was 0.70 in this study. A cut-off score of ≥8 was used to identify elevated anxiety and depression, consistent with validation studies in Thai patients with CAD [40].

Number of Rehospitalizations Questionnaire

The researcher developed the Number of Rehospitalizations Questionnaire to record unplanned rehospitalizations due to cardiac, pulmonary, or surgery-related causes without prior appointments among participants. The questionnaire consisted of two sections: unplanned hospital readmissions and acute revisit to the OPD or emergency department (ED). The CVI of the questionnaire was 1, and reliability testing was not applicable due to its event-based nature.

Data Collection

Functional status was defined as the primary outcome, with the 90-day follow-up serving as the primary endpoint, and the 30-day follow-up was included to evaluate early postoperative recovery. Anxiety, depression, and rehospitalization were assessed as secondary outcomes. Demographic and health-related data were collected using the HRDQ. Functional status was assessed using the Thai ESDS, and anxiety and depression were measured using the Thai HADS. The researcher collected baseline data on the day of admission to the Cardiac Surgical Units. Functional status, anxiety, depression, and rehospitalization were assessed 30 and 90 days after discharge or upon completion of the study. A research assistant who was blinded to group allocation collected the data via telephone interviews.

Analytical and Statistical Approaches

Data were analyzed using IBM SPSS Statistics (version 18; IBM Corp). Descriptive statistics were used to summarize demographic characteristics, health-related variables, and baseline outcomes. Prior to group comparisons of continuous (ratio-scale) variables, the assumption of normality was assessed using the Kolmogorov–Smirnov test. Categorical variables were compared using chi-square tests, while continuous variables were analyzed using 2-tailed independent *t* tests. When assumptions for the chi-square test were not met, Fisher exact test was used as an alternative.

XSL•FO
RenderX

The impact of the telehealth intervention on functional status, anxiety, and depression over time was evaluated using a 2-way repeated-measures ANOVA. This analysis considered two factors: group (intervention vs control) and time (baseline [T1], 30 days [T2], and 90 days [T3] after discharge). Repeated-measures ANOVA was selected to reduce the risk of type I error. The necessary assumptions for this analysis, including sphericity and homogeneity of variances, were verified prior to conducting the tests.

All analyses were conducted according to the intention-to-treat (ITT) principle, whereby all randomized participants were analyzed in the groups to which they were originally assigned, regardless of adherence to the intervention. Participants who discontinued the telehealth intervention (eg, failure to report measurements for five consecutive days or the occurrence of severe intraoperative or postoperative complications) were retained in the analysis in accordance with the ITT framework. Missing data for continuous outcome variables were addressed using multiple imputation with fully conditional specification. Assuming the data were missing at random, five imputed datasets were generated. Each dataset was analyzed separately, and the results were pooled using Rubin's rules to obtain the final estimates.

Differences in the proportions of participants who experienced rehospitalization due to cardiac, pulmonary, or surgery-related causes at 30 and 90 days after discharge between the intervention and control groups were evaluated using a 2-proportion *z* test. Rehospitalization was treated as a binary outcome at the participant level (event vs. no event) at each follow-up time point. Because no participants experienced repeated rehospitalization events during the follow-up period, count-based models (eg, Poisson regression) were not applicable. Therefore, a proportion-based approach using the 2-proportion *z* test was used to compare the proportion of participants experiencing rehospitalization between the intervention and control groups.

Percentages were calculated using the original randomized sample size as the denominator (n=42 per group) in accordance with the ITT principle. Participants who were lost to follow-up had no documented rehospitalization events during the monitoring period; therefore, consistent with a conservative ITT approach, these participants were classified as having experienced no rehospitalization events (coded as 0) in the analysis. Subcategories of rehospitalization events (unplanned readmission and acute care visits) were reported descriptively to illustrate the distribution of events and were not subjected to separate statistical testing. All statistical tests were two-tailed, and statistical significance was set at $P<.05$.

Ethical Considerations

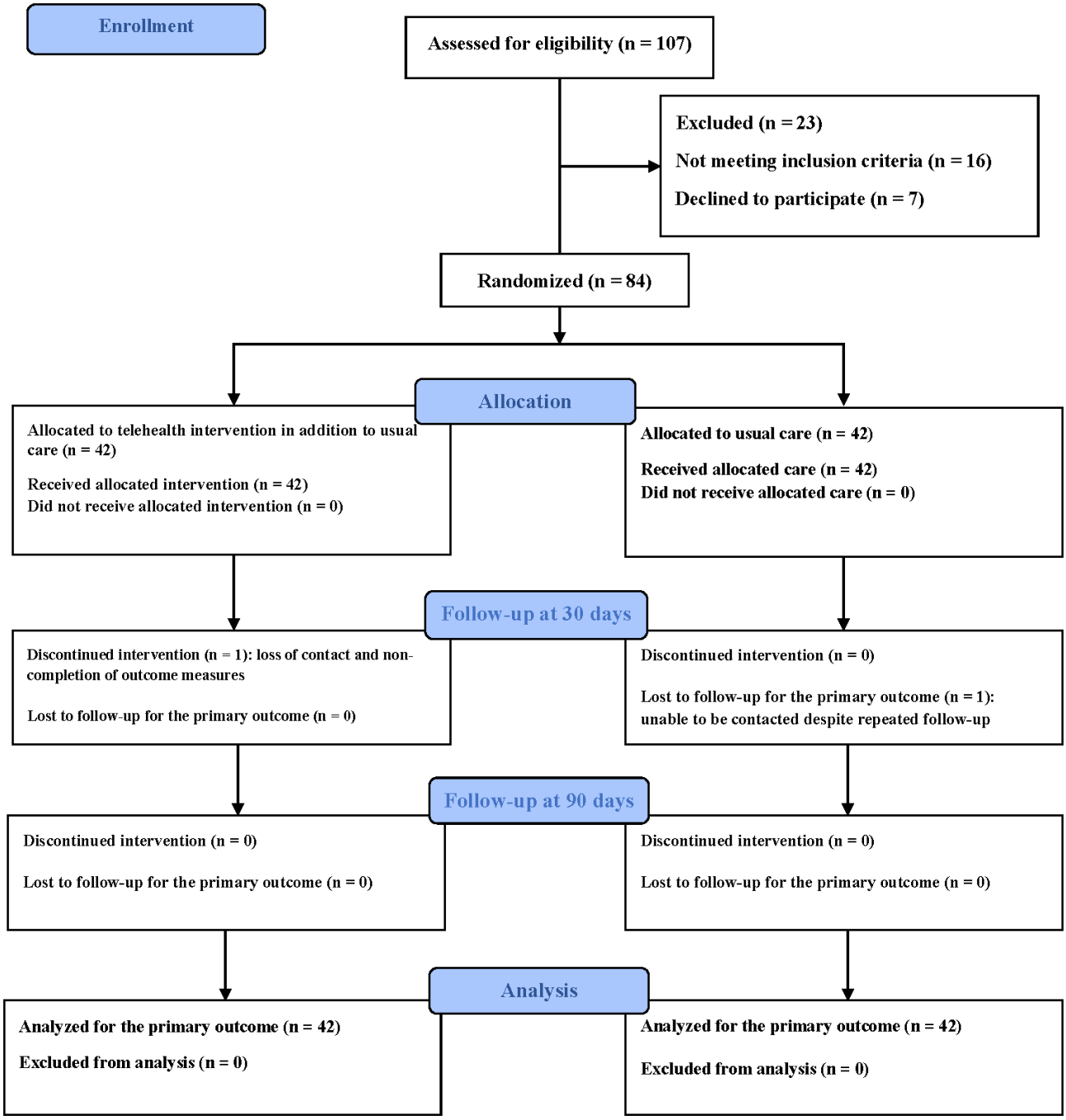
The study was reviewed and approved by the Institutional Review Board of Mahidol University (MU-MOU CoA No.IRB-NS2023/787.1807). The study was prospectively registered with the Thai Clinical Trials Registry (TCTR20230816008). All participants were provided with detailed information about the study, including the objectives, procedures, risks, benefits, confidentiality, and their right to withdraw without affecting their care. Written informed consent was obtained from all individuals prior to participation. To ensure privacy and confidentiality, all personal identifiers were removed and replaced with coded identification numbers. All data were stored securely and handled in accordance with institutional and ethical guidelines.

Results

The flow of participants through the study is presented in Figure 2. The baseline demographic and health-related characteristics of the participants were comparable between the control and intervention groups ($P>.05$; see Table 1). The mean age of the participants was 67.07 (SD 6.56) years in the intervention group and 68.86 (SD 6.83) years in the control group. Most participants in both groups were male, married, and had primary school education. Most patients were diagnosed with triple-vessel disease and had comorbidities, including hypertension, diabetes mellitus (DM), and dyslipidemia, classified as NYHA class II, and underwent on-pump CABG surgery. The average LOS in the intensive care unit was 40.36 (SD 50.42) hours in the intervention group and 47.78 (SD 40.93) hours in the control group. Similarly, the mean hospital LOS was 9.64 (SD 3.91) days in the intervention group and 11.33 (SD 6.25) days in the control group.

XSL•FO
RenderX

Figure 2. The process of the study according to the CONSORT (Consolidated Standards of Reporting Trials) flow diagram statistical analysis.



XSL•FO
RenderX

Table . Baseline demographic and health-related characteristics of participants (n=84).

Characteristics	Intervention (n=42)	Control (n=42)	*P* value
Age (years), mean (SD)	67.07 (6.56)	68.86 (6.83)	.23[a]
Male, n (%)	30 (71.4)	28 (66.7)	.64[b]
Married, n (%)	36 (85.7)	31 (73.8)	.31[b]
Primary school education, n (%)	18 (42.9)	25 (59.5)	.36[b]
Type of CAD[d], n (%)			.53[c]
DVD[e]	2 (4.8)	0 (0)	
TVD[f]	25 (59.5)	24 (57.1)	
TVD+ LM[g]	12 (28.6)	16 (38.1)	
DVD+ LM[h]	3 (7.1)	2 (4.8)	
Comorbidities, n (%)	40 (95.2)	41 (97.6)	.56[b]
HT[i]	35 (83.3)	38 (90.5)	
DM[j]	21 (50.0)	29 (69.0)	
DLP[k]	31 (73.8)	21 (50.0)	
CKD[l]	6 (14.3)	11 (26.2)	
≥2 comorbidities[m]	33 (78.6)	37 (88.1)	
NYHA[n] functional classification, n (%)			.56[b]
II	33 (78.6)	36 (85.7)	
III	9 (21.4)	6 (14.3)	
Type of CABG[o], n (%)			.31[b]
On-pump	39 (92.9)	41 (97.6)	
Off-pump	3 (7.1)	1 (2.4)	
Mean ICU[p] LOS[q], hours, mean (SD)	40.36 (50.42)	47.78 (40.93)	.46[a]
Mean Hospital LOS[q], days, mean (SD)	9.64 (3.91)	11.33 (6.25)	.14[a]

[a]Independent *t* test.

[b]Chi-square test.

[c]Fisher exact test.

[d]CAD: coronary artery disease.

[e]DVD: double-vessel disease.

[f]TVD: triple-vessel disease.

[g]TVD + LM: triple-vessel disease with left main.

[h]DVD + LM: double-vessel disease with left main.

[i]HT: hypertension.

[j]DM: diabetes mellitus.

[k]DLP: dyslipidemia.

[l]CKD: chronic kidney disease.

[m]≥2 comorbidities indicates the presence of two or more comorbid conditions (participants may have more than one comorbidity; therefore, percentages for individual conditions may exceed 100%).

[n]NYHA: New York Heart Association.

[o]CABG: coronary artery bypass grafting.

[p]ICU: intensive care unit.

[q]LOS: length of stay.

XSL•FO
RenderX

The 2-way repeated-measures ANOVA revealed a significant interaction between group and time for functional status scores ($F_{2,\ 164}$=32.09, ηp^2=.28; P<.001), anxiety ($F_{2,\ 164}$=20.22, ηp^2=.20; P<.001), and depression ($F_{2,\ 164}$=16.81, ηp^2=.17; P<.001), indicating that changes in mean scores for functional status, anxiety, and depression over time differed significantly between the intervention and control groups. Pairwise comparisons revealed significantly greater reductions in functional status scores across all time points in the intervention group than in the control group, indicating a steady improvement in functional independence. Additionally, significantly greater reductions in anxiety and depression scores were observed 30 and 90 days after discharge in the intervention group than in the control group (Table 2).

Table . Comparison of the differences in the mean scores for the study variables (n=84).

Variables and time	Intervention (n=42), mean (SD)	Control (n=42), mean (SD)	Group F (P value)	Time F (P value)	Group × time interaction F (P value)	95% CI[a]
Functional status[b]						
Baseline	23.02 (4.19)	21.17 (5.05)	15.53 (<.001)	37.63[c] (<.001)	32.09[c] (<.001)	−0.16 to 3.87
30 days	17.82 (3.82)	24.11 (6.58)	—[d]	—	—	−8.62 to −3.95
90 days	13.78 (3.12)	20.46 (7.84)	—	—	—	−9.28 to −4.10
Anxiety						
Baseline	7.94 (2.10)	7.97 (2.35)	52.17 (<.001)	84.25[c] (<.001)	20.22[c] (<.001)	−0.99 to 0.94
30 days	3.85 (1.64)	6.65 (2.87)	—	—	—	−3.81 to −1.78
90 days	1.48 (0.86)	5.74 (3.33)	—	—	—	−5.32 to −3.21
Depression						
Baseline	4.28 (1.79)	4.17 (1.77)	19.47 (<.001)	32.96[c] (<.001)	16.81[c] (<.001)	−0.67 to 0.88
30 days	2.78 (1.46)	4.38 (2.70)	—	—	—	−2.54 to −0.66
90 days	0.85 (0.94)	3.63 (2.62)	—	—	—	−3.63 to −1.92

[a]95% CI of the between-group mean difference at each time point.

[b]Lower scores of the mean scores of functional status reflect higher functional independence.

[c]Repeated-measures ANOVA sphericity assumed.

[d]Not available.

Rehospitalizations were defined as any unplanned hospital-based use occurring within 30 or 90 days after discharge, including (1) unplanned inpatient readmissions and (2) acute visits to the OPD or ED for cardiac, pulmonary, or surgery-related issues.

Within 30 days after discharge, rehospitalizations occurred significantly more frequently in the control group than in the intervention group (16/42, 38.1%) vs (5/, 11.9%); Z=2.77; P=.006). Similarly, between 31 and 90 days post discharge, rehospitalizations remained higher in the control group (5/42, 11.9%) compared with the intervention group (n=0, 0%), with a statistically significant between-group difference (Z=2.31; P=.02). Overall, participants in the intervention group experienced significantly fewer rehospitalization events than those in the control group during both follow-up intervals (Table 3).

XSL•FO
RenderX

Table . Comparison of rehospitalization events between the intervention and control groups during the 90-day follow-up period after discharge (n=84).

Variables	Intervention (n=42), n (%)[a]	Control (n=42), n (%)[a]	*Z* test[b]	*P* value
Rehospitalizations within 30 days[a]	5 (11.9)	16 (38.1)	2.77	.006
Unplanned readmission[c]	0 (0)	9 (21.4)	—[e]	—
Acute care visits[c]	5 (11.9)	7 (16.7)	—	—
Rehospitalizations between 31 and 90 days postdischarge[d]	0 (0)	5 (11.9)	2.31	.02
Unplanned readmission[c]	0 (0)	4 (9.5)	—	—
Acute care visits[c]	0 (0)	1 (2.4)	—	—

[a]Data are presented as n (%). Percentages were calculated using the intention-to-treat denominator (n=42 per group).

[b]Between-group comparisons of overall rehospitalization proportions within each time interval were conducted using the two-proportion *z* test.

[c]Subcategories of rehospitalization events (unplanned readmission and acute care visits) are presented for descriptive purposes only; no statistical comparisons were performed for these subcategories.

[d]Rehospitalization counts represent observed events occurring during the follow-up period after hospital discharge.

[e]Not available.

At 30 days after discharge, causes of unplanned hospital readmissions in the control group included heart failure (n=2), septic shock (n=2), wound infection (n=2), fatigue (n=1), hypotension (n=1), and pleural effusion (n=1). No unplanned hospital readmissions occurred in the intervention group during this period. Acute visits to the OPD or ED were also more frequent in the control group, with documented causes including hypotension (n=2), sternal wound bleeding (n=1), wound infection (n=1), severe hypertension (n=1), tiredness (n=1), and localized leg edema (n=1). In the intervention group, acute visits occurred due to wound infection (n=4) and severe pain (n=1); however, none resulted in hospital readmission.

During the 31‐90 days post discharge, unplanned hospital readmissions in the control group were attributed to heart failure (n=2), dyspnea (n=1), and wound infection (n=1). No unplanned hospital readmissions were recorded in the intervention group during this period. An acute OPD visit due to dyspnea (n=1) was documented in the control group, whereas no acute visits or rehospitalizations were observed in the intervention group. No adverse events related to the intervention were identified. Furthermore, no deaths occurred in either group during the 90-day follow-up period.

Discussion

Principal Findings

The findings of this study demonstrate the potential of a telehealth intervention, delivered through the “Zip Heart” app and video consultations, to improve functional recovery, reduce psychological symptoms, and lower rehospitalization rates among older adults undergoing CABG. Integrating telehealth into usual care appears to be a promising strategy for providing continuous support during the critical postdischarge period. The effectiveness of the intervention may be attributed to its foundation in the TCM [18], which emphasizes continuity, coordination, and sustained patient–nurse relationships—key elements for effective care management during transitions of care [17,56]. These findings are consistent with previous studies showing that TCM-based interventions reduce rehospitalization rates and health care costs while improving functional independence and quality of life in older patients following open-heart surgery [57-59].

The integration of the “Zip Heart” app with scheduled video consultations enabled effective remote monitoring and addressed both clinical and informational needs during the transition from hospital to home. Timely professional support and ongoing education facilitated a smoother recovery process. In line with these findings, prior research has shown that telehealth monitoring combined with video consultations enhances health literacy and self-management skills among older adults with chronic conditions [60], while web-based health education and telecounseling interventions are effective in alleviating postoperative psychological distress [26,61,62]. However, self-home monitoring may pose challenges for frail or vulnerable older adults due to technological barriers [28]. In this context, caregiver involvement plays a critical role in supporting recovery and facilitating successful telehealth implementation. A hybrid care model that combines telehealth with conventional in-person services may therefore be more appropriate for this population.

Although postoperative complications occurred in both study groups, the rehospitalization rate was lower in the intervention group. This suggests that the telehealth intervention may have supported earlier identification and management of postoperative issues, enabling timely outpatient treatment and reducing the need for hospital readmissions. Prompt responses to emerging complications may prevent their progression to more severe conditions requiring rehospitalization. However, rehospitalization risk is influenced by patient characteristics. In this study, participants were older adults aged 60‐88 years with a prevalence of comorbidities, including hypertension, DM, dyslipidemia, and chronic kidney disease. Socioeconomic

XSL•FO
RenderX

factors, such as low educational attainment, unemployment, and limited income, may also have contributed to reduced health literacy and difficulty understanding postoperative care instructions [63]. In addition, the absence of continuous monitoring and regular health care provider contact in the control group may have delayed the recognition of early risk factors for rehospitalization.

The findings further underscore the expanding role of nurses in delivering digitally enabled, patient-centered care. With appropriate training, nurses can effectively coordinate and manage care through telehealth platforms, thereby strengthening continuity between hospital-based and home-based services. This positions nursing practice at the forefront of implementing innovative care models tailored to the needs of an aging cardiac population. However, the telehealth intervention in this study was delivered by a specialist cardiac nurse, which may have increased personnel burden and introduced therapeutic effects related to advanced clinical expertise. The requirement for delivery by a specialist cardiac nurse represents a significant limitation, as it may constrain the generalizability and reproducibility of the intervention in settings with limited access to specialized nursing staff. Additionally, telehealth interventions play a critical role in preventing complications that lead to rehospitalization; their effectiveness may vary depending on patient characteristics, targeted outcomes, multidisciplinary involvement [64], and intervention intensity or dosage [56]. Moreover, successful implementation requires careful consideration of associated costs, including investments in technological infrastructure, staff training, and workload management strategies. Addressing these factors would be essential to ensure the long-term sustainability and scalability of telehealth-supported transitional care programs.

Limitations

Several limitations should be considered when interpreting these findings. First, this was a single-center study conducted in Thailand with a specific patient population, which may limit the generalizability of the findings to other health care settings and populations with different demographic or health system characteristics.

Second, the intervention relied on telehealth tools, including video consultations and the “Zip Heart” app. Although outcomes were assessed at 30 and 90 days post discharge, providing short- and midterm data, the long-term effects of the intervention on functional status, psychological outcomes, and rehospitalization remain unknown.

Third, the eligibility criteria required participants to have access to and be able to use a smartphone with an internet connection. This requirement may have excluded older adults with lower socioeconomic status or limited digital literacy, thereby restricting the generalizability of the findings to the broader older population.

Fourth, participants in the intervention group received more frequent contact with health care providers through scheduled video consultations compared with the control group. This increased attention may have introduced performance and attention bias (Hawthorne effect), potentially contributing to improvements in psychological outcomes. Although no participants required video consultations beyond those prespecified, the allowance for clinical discretion may still have introduced potential subjective bias. To minimize this risk, consultation frequency was aligned with the postoperative recovery trajectory, and a structured consultation guide was used to ensure consistency and reduce participant burden. In addition, the costs associated with technology implementation and staff training were not evaluated, limiting assessment of the cost-effectiveness and scalability of the intervention.

Fifth, rehospitalization data were obtained through participant or caregiver report and cross-checked against available medical records when possible. However, complete verification was not feasible for all cases due to data protection constraints, and some data relied on self-report, introducing potential recall bias. Moreover, rehospitalization was analyzed as a binary outcome, which does not capture the frequency or timing of recurrent events.

Sixth, although outcome data were collected by a blinded research assistant, the open-label design and reliance on patient-reported outcome measures may have introduced social desirability bias, potentially leading to overestimation of favorable outcomes.

Finally, 2 participants were lost to follow-up during the study period. One participant in the control group was lost to follow-up despite repeated contact attempts, and one participant in the intervention group failed to report outcome measures for five consecutive days and could not be reached by phone during the 30-day follow-up period. No additional attrition occurred during the subsequent 90-day follow-up period. The overall attrition rate was low (2.38%) and substantially lower than the anticipated attrition rate (21.4%). To maintain the integrity of randomization and minimize bias due to missing data, an ITT analysis was performed in which all randomized participants (n=84) were included in the analysis. Missing data for continuous outcomes were handled using multiple imputation, whereas a conservative zero-event assumption was applied to participants lost to follow-up for the binary outcome of rehospitalization.

Future Research Directions

This study focused exclusively on older adults undergoing CABG surgery. Future research should include other patient populations with complex health care needs to enhance the generalizability of the findings. Investigating the implementation and effectiveness of telehealth across diverse clinical contexts can broaden our understanding of its applicability and scalability. Additionally, future studies should incorporate larger sample sizes and extended follow-up periods to capture long-term outcomes. Future investigations may also consider more advanced analytical approaches to examine rehospitalization outcomes in greater detail. For example, count-based regression models (eg, Poisson or negative binomial regression) could be used to analyze the frequency of rehospitalizations when repeated events occur, while survival analysis methods could evaluate the timing of readmissions.

XSL•FO
RenderX

The telehealth intervention used in this study integrated a mobile app and video consultations to support postdischarge care. Future studies should investigate the potential of emerging technologies, such as artificial intelligence–driven platforms, wearable health-monitoring devices, and virtual reality–based rehabilitation tools. Comparative effectiveness studies evaluating these different modalities could yield important insights into identifying the most effective strategies for optimizing recovery and improving outcomes among older adults undergoing CABG.

Finally, the economic viability of telehealth interventions should be rigorously assessed through cost-effectiveness analyses, particularly with regard to their potential to reduce health care use and improve long-term patient outcomes.

Conclusions

This study demonstrated that telehealth intervention had significant positive effects on improving functional status, reducing anxiety and depression, and lowering rehospitalization rates among older patients undergoing CABG who received telehealth intervention in addition to usual care. Telehealth intervention facilitated continuous monitoring and timely support during the critical postdischarge period, thereby enhancing recovery and overall quality of care.

Acknowledgments

We gratefully acknowledge the administration, staff, and nurses of the cardiac surgical units at Siriraj Hospital, Faculty of Medicine, Mahidol University, Bangkok, Thailand, for their support and for providing the necessary resources to conduct this study. This research would not have been possible without the support of the Faculty of Graduate Studies and the Graduate Studies Alumni Association, Mahidol University. This study was funded by the Thesis Support Scholarships for Doctoral Students (Academic Year 2024), Doctor of Philosophy Program in Nursing Science (International Program), a collaborative program between the Faculty of Medicine Ramathibodi Hospital and the Faculty of Nursing, Mahidol University. We thank Associate Professor Dr. Wendy Grube, School of Nursing, University of Pennsylvania, for her assistance with English-language proofreading. Generative artificial intelligence tools, including ChatGPT (OpenAI) and Grammarly, were used to assist with language editing and manuscript refinement. In addition, native English speakers provided further review and language editing. The authors subsequently reviewed and revised the manuscript and take full responsibility for the content of the publication.

Figure 1 was partially edited using Adobe Photoshop to enhance visual clarity and presentation quality.

Funding

This research was partially supported by the Faculty of Graduate Studies and the Graduate Studies Alumni Association, Mahidol University. Additional funding was provided through the Thesis Support Scholarships for Doctoral Students (Academic Year 2024), Doctor of Philosophy Program in Nursing Science (International Program), a collaborative program between the Faculty of Medicine Ramathibodi Hospital and the Faculty of Nursing, Mahidol University, with a total grant amount of 65,000 Thai baht (approximately USD 2000).No additional external funding was received from public, commercial, or not-for-profit organizations for the research, authorship, or publication of this article.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to ethical restrictions and the need to protect participant privacy; however, they are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: JM (lead), UA (supporting), KW (supporting), PT (supporting)
Data curation: JM (lead), UA (supporting), NC (supporting)
Formal analysis: JM (lead), NC (supporting)
Funding acquisition: JM (lead), UA (supporting)
Investigation: JM (lead), UA (supporting)
Methodology: JM (lead), UA (supporting), KW (supporting), PT (supporting), NC (supporting)
Project administration: JM (lead), UA (supporting)
Supervision: UA (supporting)
Validation: JM (lead), UA (supporting), NC (supporting)
Writing – original draft: JM (lead), UA (supporting), KW (supporting), PT (supporting), NC (supporting)
Writing – review and editing: JM (lead), UA (supporting), KW (supporting), PT (supporting), NC (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theoretical framework.

XSL•FO
RenderX

[DOCX File, 14 KB - cardio_v10i1e81777_app1.docx]

Checklist 1
CONSORT checklist.
[PDF File, 230 KB - cardio_v10i1e81777_app2.pdf]

References

1. Shi H, Xia Y, Cheng Y, et al. Global burden of ischaemic heart disease from 2022 to 2050: projections of incidence, prevalence, deaths, and disability-adjusted life years. Eur Heart J Qual Care Clin Outcomes 2025 Jun 23;11(4):355-366. [doi: 10.1093/ehjqcco/qcae049] [Medline: 38918062]
2. Writing Committee Members, Lawton JS, Tamis-Holland JE, et al. ACC/AHA/SCAI guideline for coronary artery revascularization: a report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. J Am Coll Cardiol 2022 Jan 18;79(2):e21-e129. [doi: 10.1016/j.jacc.2021.09.006] [Medline: 34895950]
3. Gimbel ME, Willemsen LM, Daggelders MC, et al. Long-term follow-up after bypass surgery or coronary stenting in elderly with multivessel disease. Neth Heart J 2020 Sep;28(9):467-477. [doi: 10.1007/s12471-020-01415-z] [Medline: 32333255]
4. Açıkel MET. Evaluation of depression and anxiety in coronary artery bypass surgery patients: a prospective clinical study. Braz J Cardiovasc Surg 2019 Aug 27;34(4):389-395. [doi: 10.21470/1678-9741-2018-0426] [Medline: 31364347]
5. Correa-Rodríguez M, Abu Ejheisheh M, Suleiman-Martos N, et al. Prevalence of depression in coronary artery bypass surgery: a systematic review and meta-analysis. J Clin Med 2020 Mar 26;9(4):909. [doi: 10.3390/jcm9040909] [Medline: 32225052]
6. Mori M, Djulbegovic M, Hajduk AM, Holland ML, Krumholz HM, Chaudhry SI. Changes in functional status and health-related quality of life in older adults after surgical, interventional, or medical management of acute myocardial infarction. Semin Thorac Cardiovasc Surg 2021;33(1):72-81. [doi: 10.1053/j.semtcvs.2020.05.001] [Medline: 32439546]
7. Shawon MSR, Odutola M, Falster MO, Jorm LR. Patient and hospital factors associated with 30-day readmissions after coronary artery bypass graft (CABG) surgery: a systematic review and meta-analysis. J Cardiothorac Surg 2021 Jun 10;16(1):172. [doi: 10.1186/s13019-021-01556-1] [Medline: 34112216]
8. Zea-Vera R, Zhang Q, Amin A, et al. Development of a risk score to predict 90-day readmission after coronary artery bypass graft. Ann Thorac Surg 2021 Feb;111(2):488-494. [doi: 10.1016/j.athoracsur.2020.04.142] [Medline: 32585200]
9. Thai society of thoracic surgeons [Web page in Thai]. The Society of Thoracic Surgeons of Thailand. URL: https://www.thaists.or.th [accessed 2024-03-07]
10. Manapunsopee S, Thanakiatpinyo T, Wongkornrat W, Chuaychoo B, Thirapatarapong W. Effectiveness of incentive spirometry on inspiratory muscle strength after coronary artery bypass graft surgery. Heart Lung Circ 2020 Aug;29(8):1180-1186. [doi: 10.1016/j.hlc.2019.09.009] [Medline: 31735684]
11. Sanonoi N. Effects of a telehealth program on the functional status and rehospitalization rates of patients after coronary artery bypass grafts. Chula Med 2021;65(2):179-186. [doi: 10.58837/CHULA.CMJ.65.2.10]
12. Niemeyer-Guimarães M, Cendoroglo MS, Almada-Filho CM. Course of functional status in elderly patients after coronary artery bypass surgery: 6-month follow up. Geriatr Gerontol Int 2016 Jun;16(6):737-746. [doi: 10.1111/ggi.12547] [Medline: 26178448]
13. Mortensen M, Sandvik R, Svendsen Ø, Haaverstad R, Moi AL. Return to work after coronary artery bypass grafting and aortic valve replacement surgery: a scoping review. Scand J Caring Sci 2022 Dec;36(4):893-909. [doi: 10.1111/scs.13006] [Medline: 34057755]
14. Tully PJ, Winefield HR, Baker RA, et al. Depression, anxiety and major adverse cardiovascular and cerebrovascular events in patients following coronary artery bypass graft surgery: a five year longitudinal cohort study. Biopsychosoc Med 2015;9:14. [doi: 10.1186/s13030-015-0041-5] [Medline: 26019721]
15. Bjørnnes AK, Parry M, Lie I, et al. The impact of an educational pain management booklet intervention on postoperative pain control after cardiac surgery. Eur J Cardiovasc Nurs 2017 Jan;16(1):18-27. [doi: 10.1177/1474515116631680] [Medline: 26846145]
16. Singh Solorzano C, Rowlands H, Ronaldson A, et al. Caregiver burden in informal spousal caregivers predicts psychological and physical health in patients following coronary artery bypass graft surgery: a longitudinal clinical cohort study. J Behav Med 2024 Dec;47(6):1118-1126. [doi: 10.1007/s10865-024-00512-3] [Medline: 39183251]
17. Hirschman KB, Shaid E, McCauley K, Pauly MV, Naylor MD. Continuity of care: the transitional care model. Online J Issues Nurs 2015 Sep 30;20(3):1. [Medline: 26882510]
18. Naylor MD. Advancing high value transitional care: the central role of nursing and its leadership. Nurs Adm Q 2012;36(2):115-126. [doi: 10.1097/NAQ.0b013e31824a040b] [Medline: 22407204]
19. Sawalha O, Ariza-Vega P, Alhalaiqa F, Pérez-Rodríguez S, Romero-Ayuso D. Psychological discomfort in patients undergoing coronary artery bypass graft (CABG) in West Bank: a cohort study. J Clin Med 2024 Mar 30;13(7):2027. [doi: 10.3390/jcm13072027] [Medline: 38610792]

XSL•FO
RenderX

20. Wu KA, Kunte S, Rajkumar S, et al. Digital health for patients undergoing cardiac surgery: a systematic review. Healthcare (Basel) 2023 Aug 28;11(17):2411. [doi: 10.3390/healthcare11172411] [Medline: 37685445]
21. Awaludin S, Nurachmah E, Soetisna TW, Umar J. The effect of a smartphone-based perioperative nursing intervention: prayer, education, exercise therapy, hypnosis, and music toward pain, anxiety, and early mobilization on cardiac surgery. J Public Health Res 2021 Dec 2;11(2):2742. [doi: 10.4081/jphr.2021.2742] [Medline: 35255671]
22. Champion S, Clark RA, Tirimacco R, Tideman P, Gebremichael L, Beleigoli A. The impact of the SARS-CoV-2 Virus (COVID-19) pandemic and the rapid adoption of telehealth for cardiac rehabilitation and secondary prevention programs in rural and remote Australia: a multi-method study. Heart Lung Circ 2022 Nov;31(11):1504-1512. [doi: 10.1016/j.hlc.2022.07.006] [Medline: 35987722]
23. Chao GF, Li KY, Zhu Z, et al. Use of telehealth by surgical specialties during the COVID-19 pandemic. JAMA Surg 2021 Jul 1;156(7):620-626. [doi: 10.1001/jamasurg.2021.0979] [Medline: 33769434]
24. Dalal HM, Doherty P, McDonagh ST, Paul K, Taylor RS. Virtual and in-person cardiac rehabilitation. BMJ 2021 Jun 3;373:n1270. [doi: 10.1136/bmj.n1270] [Medline: 34083376]
25. Kleinpell RM, Avitall B, Catrambone C, Johnson T, Fogg L. Randomized trial of a discharge planning and telehealth intervention for patients aged 65 and older after coronary artery bypass surgery. Int J Clin Cardiol 2015;2(4) [FREE Full text] [doi: 10.23937/2378-2951/1410044]
26. Li J, Deng Y, Jiang Y. The effectiveness of a web-based information-knowledge-attitude-practice continuous intervention on the psychological status, medical compliance, and quality of life of patients after coronary artery bypass grafting surgery: a parallel randomized clinical trial. J Cardiothorac Surg 2024 Mar 13;19(1):125. [doi: 10.1186/s13019-024-02618-w] [Medline: 38481263]
27. Pakrad F, Ahmadi F, Grace SL, Oshvandi K, Kazemnejad A. Traditional vs extended hybrid cardiac rehabilitation based on the continuous care model for patients who have undergone coronary artery bypass surgery in a middle-income country: a randomized controlled trial. Arch Phys Med Rehabil 2021 Nov;102(11):2091-2101. [doi: 10.1016/j.apmr.2021.04.026] [Medline: 34175270]
28. Sarkar S, MacLeod J, Hassan A, Brunt KR, Palmer K, Légaré JF. Enhanced telehealth home-monitoring intervention for vulnerable and frail patients after cardiac surgery (THE-FACS pilot intervention study). BMC Geriatr 2022 Nov 5;22(1):836. [doi: 10.1186/s12877-022-03531-4] [Medline: 36333652]
29. van Steenbergen G, van Veghel D, van Lieshout D, Sperwer M, Ter Woorst J, Dekker L. Effects of video-based patient education and consultation on unplanned health care utilization and early recovery after coronary artery bypass surgery (IMPROV-ED): randomized controlled trial. J Med Internet Res 2022 Aug 26;24(8):e37728. [doi: 10.2196/37728] [Medline: 36018625]
30. Wu M, Li C, Hu T, et al. Effectiveness of telecare interventions on depression symptoms among older adults: systematic review and meta-analysis. JMIR Mhealth Uhealth 2024 Jan 17;12:e50787. [doi: 10.2196/50787] [Medline: 38231546]
31. Bostrom J, Sweeney G, Whiteson J, Dodson JA. Mobile health and cardiac rehabilitation in older adults. Clin Cardiol 2020 Feb;43(2):118-126. [doi: 10.1002/clc.23306] [Medline: 31825132]
32. Jamal NE, Abi-Saleh B, Isma'eel H. Advances in telemedicine for the management of the elderly cardiac patient. J Geriatr Cardiol 2021 Sep 28;18(9):759-767. [doi: 10.11909/j.issn.1671-5411.2021.09.004] [Medline: 34659382]
33. Noel K, Messina C, Hou W, Schoenfeld E, Kelly G. Tele-transitions of care (TTOC): a 12-month, randomized controlled trial evaluating the use of telehealth to achieve triple aim objectives. BMC Fam Pract 2020 Feb 7;21(1):27. [doi: 10.1186/s12875-020-1094-5] [Medline: 32033535]
34. Qi K, Koike T, Yasuda Y, Tayama S, Wati I. The effects on rehospitalization rate of transitional care using information communication technology in patients with heart failure: a scoping review. Int J Nurs Stud Adv 2023 Dec;5:100151. [doi: 10.1016/j.ijnsa.2023.100151] [Medline: 38746554]
35. Gonçalves-Bradley DC, J Maria AR, Ricci-Cabello I, et al. Mobile technologies to support healthcare provider to healthcare provider communication and management of care. Cochrane Database Syst Rev 2020;2020(8):CD012927. [doi: 10.1002/14651858.CD012927.pub2]
36. Totten AM, Womack DM, Eden KB, et al. Telehealth: Mapping the Evidence for Patient Outcomes From Systematic Reviews: Rockville (MD): Agency for Healthcare Research and Quality (US); 2016. [Medline: 27536752]
37. Lai CY, Lin CH, Chao TC, et al. Effectiveness of a 12-week telerehabilitation training in people with long COVID: a randomized controlled trial. Ann Phys Rehabil Med 2024 Jun;67(5):101853. [doi: 10.1016/j.rehab.2024.101853] [Medline: 38824899]
38. Chiang SL, Shen CL, Chen LC, Lo YP, Lin CH, Lin CH. Effectiveness of a home-based telehealth exercise training program for patients with cardiometabolic multimorbidity: a randomized controlled trial. J Cardiovasc Nurs 2020;35(5):491-501. [doi: 10.1097/JCN.0000000000000693] [Medline: 32511110]
39. Chiang SL, Shen CL, Lee MS, Lin CH, Lin CH. Effectiveness of a 12-week tele-exercise training program on cardiorespiratory fitness and heart rate recovery in patients with cardiometabolic multimorbidity. Worldviews Evid Based Nurs 2023 Aug;20(4):339-350. [doi: 10.1111/wvn.12607] [Medline: 36193903]

XSL•FO
RenderX

40. Yuroong A, Asdornwised U, Pinyopasakul W, Wongkornrat W, Chansatitporn N. The effectiveness of the transitional care program among people awaiting coronary artery bypass graft surgery: a randomized control trial. J Nurs Scholarsh 2021 Sep;53(5):585-594. [doi: 10.1111/jnu.12673] [Medline: 34013579]
41. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
42. Masterson Creber R, Dodson JA, Bidwell J, et al. Telehealth and health equity in older adults with heart failure: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2023 Nov;16(11):e000123. [doi: 10.1161/HCQ.0000000000000123] [Medline: 37909212]
43. Liu N, Yin J, Tan SSL, Ngiam KY, Teo HH. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021 Oct 12;28(11):2483-2501. [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
44. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
45. Keeping-Burke L, Purden M, Frasure-Smith N, Cossette S, McCarthy F, Amsel R. Bridging the transition from hospital to home: effects of the VITAL telehealth program on recovery for CABG surgery patients and their caregivers. Res Nurs Health 2013 Dec;36(6):540-553. [doi: 10.1002/nur.21571] [Medline: 24242195]
46. Ritklar L. Results of discharge planning by using D-METHOD on re-admission and satisfaction in patients with coronary artery disease. TUHJ 2018;3(3):19-27 [FREE Full text]
47. Twisk JW. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide, 2nd edition: Cambridge University Press; 2013. URL: https://www.cambridge.org/core/product/identifier/9781139342834/type/book [doi: 10.1017/CBO9781139342834]
48. Comprehensive discharge planning and home follow up of Thai elderly coronary artery bypass graft (CABG) patients [Report in Thai]. : Thailand Science Research and Innovation, Thai Research Fund Report; 2006 URL: https://elibrary.tsri.or.th/fullP/PDF4480038/PDF4480038_full.pdf [accessed 2026-04-25]
49. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
50. The classification of cardiac diagnosis. JAMA 1921 Oct 29;77(18):1414. [doi: 10.1001/jama.1921.02630440034013]
51. Trongsakul S, Lambert R, Clark A, Wongpakaran N, Cross J. Development of the Thai version of Mini-Cog, a brief cognitive screening test. Geriatr Gerontol Int 2015 May;15(5):594-600. [doi: 10.1111/ggi.12318] [Medline: 25155468]
52. Benoliel JQ, McCorkle R, Young K. Development of a social dependency scale. Res Nurs Health 1980 Mar;3(1):3-10. [doi: 10.1002/nur.4770030103] [Medline: 6901191]
53. McCorkle R, Benoliel JQ, Mumma CM, Germino BB, Georgiadou F, Driever MJ. Cancer patient responses to psychosocial variables. West J Nurs Res 1982 Aug;4(3):133-141. [doi: 10.1177/019394598200400325]
54. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
55. Nilchaikovit T. Development of Thai version of hospital anxiety and depression scale in cancer patients. J Psychiatr Assoc Thai 1996;4:0125-6985 [FREE Full text]
56. Morkisch N, Upegui-Arango LD, Cardona MI, et al. Components of the transitional care model (TCM) to reduce readmission in geriatric patients: a systematic review. BMC Geriatr 2020 Sep 11;20(1):345. [doi: 10.1186/s12877-020-01747-w] [Medline: 32917145]
57. Coskun S, Duygulu S. The effects of nurse led transitional care model on elderly patients undergoing open heart surgery: a randomized controlled trial. Eur J Cardiovasc Nurs 2022 Jan 11;21(1):46-55. [doi: 10.1093/eurjcn/zvab005] [Medline: 33821999]
58. Dığın F, Fındık Ü. The effect of nurse telephone consultation after coronary artery bypass on the autonomy level of elderly patients: a quasi-experimental study. Med J Bakirkoy 2021 Jul 1;17(2):135-141. [doi: 10.4274/BMJ.galenos.2021.19970]
59. Jepma P, Verweij L, Buurman BM, et al. The nurse-coordinated cardiac care bridge transitional care programme: a randomised clinical trial. Age Ageing 2021 Nov 10;50(6):2105-2115. [doi: 10.1093/ageing/afab146] [Medline: 34304264]
60. Banbury A, Nancarrow S, Dart J, et al. Adding value to remote monitoring: co-design of a health literacy intervention for older people with chronic disease delivered by telehealth - the telehealth literacy project. Patient Educ Couns 2020 Mar;103(3):597-606. [doi: 10.1016/j.pec.2019.10.005] [Medline: 31744701]
61. Noor Hanita Z, Khatijah LA, Kamaruzzaman S, Karuthan C, Raja Mokhtar RA. A pilot study on development and feasibility of the “MyEducation: CABG application” for patients undergoing coronary artery bypass graft (CABG) surgery. BMC Nurs 2022 Feb 4;21(1):40. [doi: 10.1186/s12912-022-00814-4] [Medline: 35120517]
62. De Cola MC, Maresca G, D’Aleo G, et al. Teleassistance for frail elderly people: a usability and customer satisfaction study. Geriatr Nurs 2020;41(4):463-467. [doi: 10.1016/j.gerinurse.2020.01.019] [Medline: 32067831]
63. Dimech N, Cassar M, Carabott J. The health literacy and patient discharge experience dyad after cardiac surgery: an exploratory study. Br J Nurs 2023 Nov 23;32(21):1046-1052. [doi: 10.12968/bjon.2023.32.21.1046] [Medline: 38006591]

XSL•FO
RenderX

64. Sezgin D, O'Caoimh R, Liew A, et al. The effectiveness of intermediate care including transitional care interventions on function, healthcare utilisation and costs: a scoping review. Eur Geriatr Med 2020 Dec;11(6):961-974. [doi: 10.1007/s41999-020-00365-4] [Medline: 32754841]

Abbreviations

CABG: coronary artery bypass grafting
CAD: coronary artery disease
CONSORT: Consolidated Standards of Reporting Trials
CVI: content validity index
D-METHOD: diagnosis, medications, environment, treatment, health behaviors, outpatient follow-up, and diet
DM: diabetes mellitus
ED: emergency department
ESDS: Enforced Social Dependency Scale
HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression scale–Anxiety Subscale
HADS-D: Hospital Anxiety and Depression Scale–Depression Subscale
HRDQ: Health-Related Data Questionnaire
ITT: intention-to-treat
LOS: length of stay
NYHA: New York Heart Association
OPD: outpatient department
TCM: transitional care model

Edited by A Coristine; submitted 03.Aug.2025; peer-reviewed by GV Steenbergen, NJ Patil; revised version received 13.Apr.2026; accepted 13.Apr.2026; published 07.May.2026.

Please cite as:
Mala J, Asdornwised U, Wongkongkam K, Chansatitporn N, Thongcharoen P
Effectiveness of a Telehealth Intervention on Functional Status, Anxiety, Depression, and Rehospitalization Among Older Adults Undergoing Coronary Artery Bypass Grafting: Randomized Controlled Trial
JMIR Cardio 2026;10:e81777
URL: https://cardio.jmir.org/2026/1/e81777
doi:10.2196/81777

© Jirawan Mala, Usavadee Asdornwised, Kessiri Wongkongkam, Natkamol Chansatitporn, Punnarerk Thongcharoen. Originally published in JMIR Cardio (https://cardio.jmir.org), 7.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Influence of Perioperative Variables on Surgical Site Infection in Patients With Frailty and Prefrailty Undergoing Coronary Artery Bypass Grafting: Retrospective Multicenter Study

Mauricio Guerrieri[1], PhD; Omar Mejia[1], PhD; Bianca Meneghini[1], PhD; Natali Rodrigues[2], MD; William Burato[2], MD; Maisa Barros[2], MD; Daniel Wernsbach[2], MD; Lucas Bartelega[2], MD; Alexandre Hueb[2], PhD; Fabio Jatene[1], PhD

[1]Medical School, Universidade de São Paulo, Av. Dr. Enéas Carvalho de Aguiar, 44, Cerqueira César, São Paulo, Brazil

[2]Medical School, Universidade do Vale do Sapucaí, Pouso Alegre, Brazil

Corresponding Author:
Mauricio Guerrieri, PhD
Medical School, Universidade de São Paulo, Av. Dr. Enéas Carvalho de Aguiar, 44, Cerqueira César, São Paulo, Brazil

Abstract

Background: Surgical site infection (SSI) in patients with frailty undergoing myocardial revascularization is a major clinical challenge, and there is a gap in the literature related to this topic.

Objective: This study aimed to identify perioperative variables associated with SSI in patients with frailty and prefrailty undergoing coronary artery bypass grafting within 30 days of the procedure.

Methods: This retrospective observational study included 467 patients with frailty or prefrailty aged 60 years and older who underwent coronary artery bypass grafting in 6 Brazilian centers between January 2018 and December 2024. Frailty was defined using the Fried criteria. SSI was defined as a wound with exposed tissue with positive culture requiring antibiotic treatment. Univariate and multivariate analyses were carried out.

Results: The rate of SSI was 5.6% (26/467). The New York Heart Association classification (P=.02), Charlson Comorbidity Index (CCI; P=.01), use of cardiopulmonary bypass (CPB; P=.003), CPB time (P=.02), total surgery time (P<.001), occurrence of bleeding (P=.04), mechanical ventilation time (P=.001), length of hospital stay (P<.001), intensive care unit stay (P<.001), and need for reoperation (P<.001) variables were associated with SSI in unadjusted comparisons. In the multivariate model, the CCI (odds ratio 1.58, 95% CI 1.25 - 1.99; P<.001) and use of CPB (odds ratio 4.89, 95% CI 1.71 - 13.94; P=.003) were independently associated with SSI.

Conclusions: The CCI and the use of CPB were independently associated with SSI.

(JMIR Cardio 2026;10:e79408) doi:10.2196/79408

KEYWORDS

coronary disease; coronary artery bypass; surgical wound infection; aged; frailty

Introduction

Longevity is increasing in many countries, where life expectancy at birth is rising [1]. A reduction in mortality among older age groups seems to be one of the main reasons for this increase [2], and consequently, living with chronic, incurable diseases is becoming more frequent [3]. Population aging, improved public health policies, and advances in cardiovascular disease prevention and treatment have increased the number of people living with chronic disease because more individuals survive acute cardiovascular events, increasing the prevalence and absolute number of older adults living with these conditions and their complications [4]. This compromises the quality of life of these individuals and highlights the concept of frailty, which is gaining prominence in clinical practice [5]. Frailty is the physiological decline of the body's functional reserves [5,6].

Aging-associated frailty increases the risk of postoperative complications, and age and frailty are recognized as risk factors for mortality and morbidity in patients undergoing surgical procedures [5,7,8]. Coronary artery bypass grafting (CABG) is a well-established, common procedure; specifically, 169,705 operations were performed in 2020 in the United States and 19,125 procedures in Brazil's Unified Health System in 2023 [9,10]. Given these figures, it is essential to consider the morbidity associated with this procedure and seek better clinical outcomes and quality of life for patients.

Among the postoperative complications that arise within 30 days, surgical site infection (SSI) stands out for its clinical

XSL•FO
RenderX

relevance, occurring in 0.3% to 3.5% of myocardial revascularization cases [11-16]. The associated mortality rate varies from 10% to 47% in cases of SSI, significantly contributing to an increase in morbidity, mortality, and hospital costs and potentially trebling the procedure's cost and having a considerable impact on public resources [15,17-19].

SSI is associated with worse short- and long-term clinical outcomes. Interestingly, age alone does not seem to be a relevant predictive factor for this outcome [20-22]. Moreover, frailty has not been properly investigated as a possible risk factor for SSI in patients undergoing myocardial revascularization.

It is well established that frailty increases morbidity and mortality in older adults as well as the risk of postoperative complications by up to 2.54 times [5,7]. Therefore, it is plausible that individuals with frailty may be more susceptible to SSI than the general population. There may be additional factors associated with this potential increase in risk, which, if identified, will facilitate the development of specific preventive strategies for this patient population.

The objective of this study was to identify the perioperative variables associated with SSI in patients with frailty and prefrailty undergoing CABG within 30 days after surgery.

Methods

Study Design and Participants

This was a retrospective, longitudinal secondary analysis of data from the multicenter randomized FRAGILE (Off-Pump Versus On-Pump Coronary Artery Bypass Grafting in Frail Patients) trial, in which enrolled participants were randomly assigned to undergo either on-pump (cardiopulmonary bypass; CPB) or off-pump CABG surgery in 6 Brazilian hospitals [23]. Data were collected using the FRAGILE study dataset with standardized variable definitions across centers; however, some inter-institutional variation in clinical practice and documentation cannot be excluded. In the FRAGILE study, CPB use was assigned through randomization rather than surgeon preference; therefore, CPB was not subject to confounding by indication at the point of treatment allocation.

Although the FRAGILE protocol initially planned for the participation of 10 centers, the final dataset comprised 6 centers due to incomplete data capture and reduced recruitment at some sites during the COVID-19 pandemic period. The present secondary analysis used the final FRAGILE dataset to preserve data completeness and integrity. The data were extracted in March 2025 after the research ethics committee approved the secondary analysis and waived the requirement to obtain new informed consent from the participants.

This study included data from 467 patients aged 60 years and older with 2 or more Fried frailty criteria who underwent on- or off-pump CABG between January 2018 and December 2024 [5]. As a retrospective secondary analysis of an existing multicenter study, the sample size was determined by data availability over the study period rather than an a priori calculation. Patients who underwent urgent and emergency surgery and combined cardiac surgery, those with creatinine clearance below 30 mL per minute and hemoglobin A_{1c} above 8.5%, and those with a history of cardiac surgery were excluded because these conditions are strong determinants of infectious complications and perioperative morbidity and could introduce substantial confounding. Missing data were minimal for the variables used in the final model; therefore, a complete-case analysis was performed. Frailty was assessed based on the Fried criteria, which include unintended weight loss, self-reported exhaustion, weakness, slowness, and low physical activity. Patients with 2 criteria were considered prefrail, whereas those with 3 or more criteria were deemed to be frail. We included both patients with prefrailty and frailty because the FRAGILE study defined vulnerability using 2 or more Fried criteria. We acknowledge that inclusion of participants with prefrailty may reduce discrimination compared with a frail-only cohort and interpreted the frailty score accordingly.

SSI was operationally defined using predefined criteria captured in the FRAGILE dataset aligned with Society of Thoracic Surgeons–related documentation of clinically significant sternal wound complications:

- An open wound with exposure of muscle or subcutaneous tissue or need for reoperation due to infection [24]
- Positive culture unless the patient was taking antibiotics at the time of sample collection for bacterial culture [24]
- Treatment with antibiotics in addition to perioperative prophylaxis [24]

The primary outcome was the occurrence of any SSI within 30 days after CABG. All recorded events—including sternal dehiscence, superficial SSI, deep SSI, and mediastinitis—were combined into a single composite SSI outcome. Each event was also classified by subtype for descriptive reporting only. Subtype-specific analyses were not performed because of the limited number of events in each category.

The preoperative variables analyzed were age, sex, BMI, diabetes mellitus, creatinine clearance, New York Heart Association (NYHA) functional class, Charlson Comorbidity Index (CCI), and frailty score. Intraoperative variables included the number of mammary arteries used, CPB use, and CPB and surgery duration. The secondary outcomes included postoperative variables such as occurrence of bleeding, duration of mechanical ventilation, duration of hospitalization, and duration of intensive care unit (ICU) stay. SSI was assessed within 30 days after surgery. Length of hospital stay refers to the total hospitalization time of patients who developed SSI, including the period after SSI diagnosis and subsequent treatment during the same hospitalization. Potential collinearity between CPB time and surgery duration was considered during model building; to preserve model parsimony and avoid redundancy, only the clinically most informative variables were retained in the final adjusted model.

The CCI is a weighted index comprising 19 comorbid conditions and was originally developed to estimate mortality risk [25]. Higher scores indicate a greater comorbidity burden. In this study, the CCI was analyzed as a continuous variable.

Postoperative variables were included for exploratory association analysis and should not be interpreted as

XSL•FO
RenderX

preoperative risk predictors. This analysis was performed with the aim of understanding the distribution, trends, and patterns of each variable in isolation and in relation to the occurrence of SSI. Although additional intraoperative complexity variables were available in the FRAGILE dataset, including left ventricular ejection fraction, number of distal anastomoses, bilateral mammary artery use, CPB time, and cross-clamp time, these variables were not all included in the final multivariable model because of the limited number of SSI events. To reduce the risk of overfitting and unstable estimates, the final model was kept parsimonious and focused on variables with stronger statistical and clinical relevance.

Statistical Analysis

Statistical analyses compared baseline and perioperative characteristics between patients with and without SSI within 30 days after CABG. Secondary postoperative outcomes (bleeding, mechanical ventilation duration, ICU stay, and hospital length of stay) were also compared between groups. The continuous variables were initially assessed for normal distribution using the Shapiro-Wilk test and expressed as medians and IQRs owing to the asymmetrical distribution of the data. The Mann-Whitney *U* test for nonparametric data was used to compare 2 independent groups: occurrence of infection and no infection.

The categorical variables were expressed as absolute and relative values and compared between groups using the Pearson chi-square test. The Fisher exact test was used in the presence of expected frequencies of less than 5 in one or more cells of the contingency table. In the case of categorical variables with multiple ordinal categories, such as the number of mammary arteries used, the likelihood ratio test was used. Univariable screening may fail to retain important confounders that are not individually significant, which may bias adjusted estimates in sparse-event settings. Additionally, multiple logistic regression analysis was performed to identify the factors independently associated with the occurrence of postoperative outcomes. The variables with statistical significance in the univariate analysis ($P<.05$) and those considered clinically relevant were included in the model. The results were expressed as odds ratios (ORs) with their corresponding 95% CIs and *P* values. Analyses were performed according to randomized treatment allocation using CPB assignment as the exposure. The level of statistical significance set for all analyses was 5% ($P<.05$), and the analyses were carried out using the SPSS software (version 19; IBM Corp).

Ethical Considerations

The study protocol was approved by the Research Ethics Committee of Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, Brazil (Certificate of Presentation for Ethical Consideration 30702020.0.0000.0068; approval number 4.228.709; approved on August 21, 2020). This study was conducted in accordance with Brazilian National Health Council Resolution 466/2012 and the ethical principles of the Declaration of Helsinki.

Because this was a retrospective study based on the analysis of previously collected clinical data and medical records, the requirement for individual informed consent was waived by the Research Ethics Committee. All data used in the study were coded and deidentified before analysis. Access to the research database was restricted to the authorized research team, and no information capable of identifying individual participants is presented in the manuscript. Participants received no financial or other compensation for inclusion in this study.

Results

The sample consisted of 467 patients, 26 (5.6%) of whom had SSI. The classification of the type of SSI was as follows: 1.9% (n=9) sternal dehiscence, 1.5% (n=7) superficial SSI, 1.1% (n=5) deep SSI, 0.4% (n=2) deep infection with use of corticosteroids, 0.4% (n=2) mediastinitis, and 0.2% (n=1) subtype not specified in the database (Table 1). The classification of “deep infection with corticosteroid use” reflects the original FRAGILE dataset subtype classification; subtype categories were reported descriptively only and were not used in the univariable or multivariable models.

Table 2 shows the results related to the preoperative variables associated with SSI.

The median age was similar between the groups with and without infection (70, IQR 65 - 74 vs 69, IQR 65 - 74 years, respectively; $P=.79$), as was the median BMI (28, IQR 24 - 33 vs 26, IQR 24 - 29 kg/m^2, respectively; $P=.07$). Between the 2 groups, there were no statistically significant differences in sex (male: 14/26, 53.8% vs 303/441, 68.7%, respectively; $P=.12$), presence of diabetes mellitus (10/26, 38.5% vs 153/441, 34.7%, respectively; $P=.70$), or creatinine clearance (median 63, IQR 48 - 82 vs 62, IQR 51 - 76 mL/min, respectively; $P=.91$).

Table . Classification of types of surgical site infection (N=467).

Type of surgical site infection	Events, n (%)
Sternal dehiscence	9 (1.9)
Superficial surgical site infection	7 (1.5)
Deep surgical site infection	5 (1.1)
Deep surgical site infection with corticosteroid use	2 (0.4)
Mediastinitis	2 (0.4)
Subtype not specified	1 (0.2)

XSL•FO
RenderX

Table . Preoperative variables in patients with and without surgical site infection (SSI).

Variables	Infection		*P* value
	No (n=441)	Yes (n=26)	
Age (y), median (IQR)	69 (65-74)	70 (65-74)	.79[a]
Sex (male), n (%)	303 (68.7)	14 (53.8)	.12[b]
BMI (kg/m^2), median (IQR)	26 (24-29)	28 (24-33)	.07[a]
Diabetes mellitus, n (%)	153 (34.7)	10 (38.5)	.70[b]
Creatinine clearance (mL/min), median (IQR)	62 (51-76)	63 (48-82)	.91[a]
NYHA[c] class, n (%)			.02[d]
I	184 (41.7)	8 (30.8)	
II	149 (33.8)	6 (23.1)	
III	77 (17.5)	5 (19.2)	
IV	31 (7.0)	7 (26.9)	
Charlson Comorbidity Index, median (IQR)	4 (3-5)	5 (4-6)	.01[a]
Frailty score, median (IQR)	4 (3-4)	4 (3-5)	.49[a]

[a]Mann-Whitney *U* test.

[b]Pearson chi-square test.

[c]NYHA: New York Heart Association.

[d]Likelihood ratio test.

However, there was a significant association between NYHA functional class and the occurrence of infection (*P*=.02), with a higher proportion of patients in class IV in the group with infection (7/26, 26.9% vs 31/441, 7.0%; Figure 1). The CCI was higher among patients with infection (median 5, IQR 4 - 6 vs 4, IQR 3 - 5; *P*=.01; Figure 2). The frailty score was slightly higher in the infection group than in the group without infection but without statistical significance (median 4, IQR 3 - 5 vs 4, IQR 3 - 4, respectively; *P*=.49). Table 3 shows the results obtained for the intraoperative variables.

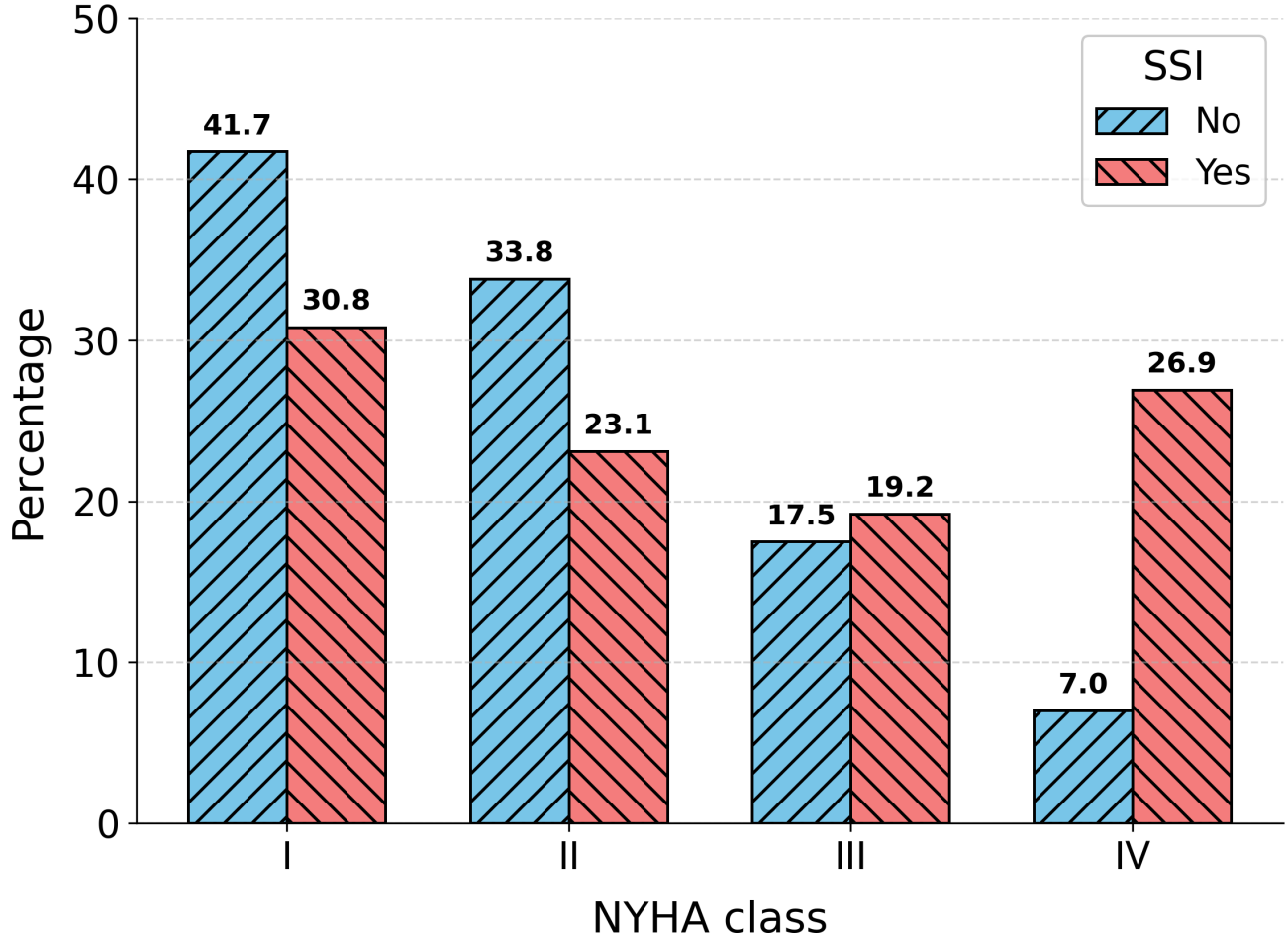
The number of mammary arteries used also showed no statistically significant difference between the infection and no infection groups (*P*=.52; ie, using only 1 mammary artery [20/26, 76.9% vs 298/441, 67.6%, respectively] or using 2 mammary arteries [6/26, 23.1% vs 133/441, 30.2%, respectively]). The use of CPB was more frequent in the group with SSI (21/26, 80.8% vs 223/441, 50.6%; *P*=.003; Figure 3). CPB time was statistically associated with SSI in univariable analysis (median 85, IQR 71 - 128 vs 85, IQR 65 - 105 minutes in the infection and no infection groups, respectively; *P*=.02). Although the medians were identical, the distribution differed between groups (IQR shift), and this finding should be interpreted cautiously. Additionally, the total time of the surgical procedure was different between the groups, with prolonged surgery time being associated with a higher occurrence of infection (median 303, IQR 250 - 359 vs 270, IQR 217 - 325 minutes in the infection and no infection groups, respectively; *P*<.001). Table 4 shows the results obtained for the postoperative variables.

Occurrence of bleeding was higher in the group with infection (2/26, 7.7% vs 4/441, 0.9%; *P*=.04). The duration of mechanical ventilation was significantly longer among patients with infection (median 18, IQR 8 - 21 vs 8, IQR 5 - 14 hours; *P*=.001), as were the total length of hospital stay (median 41, IQR 23 - 55 vs 7, IQR 6 - 11 days; *P*<.001) and the length of ICU stay (median 5, IQR 4 - 15 vs 3, IQR 2 - 5 days; *P*<.001). The need for reoperation was considerably more frequent in the group with infection (13/26, 50% vs 13/441, 2.9%; *P*<.001).

In the multiple logistic regression analysis, 2 variables were identified as independently associated with SSI (Table 5). All 467 patients were included in the multivariable regression.

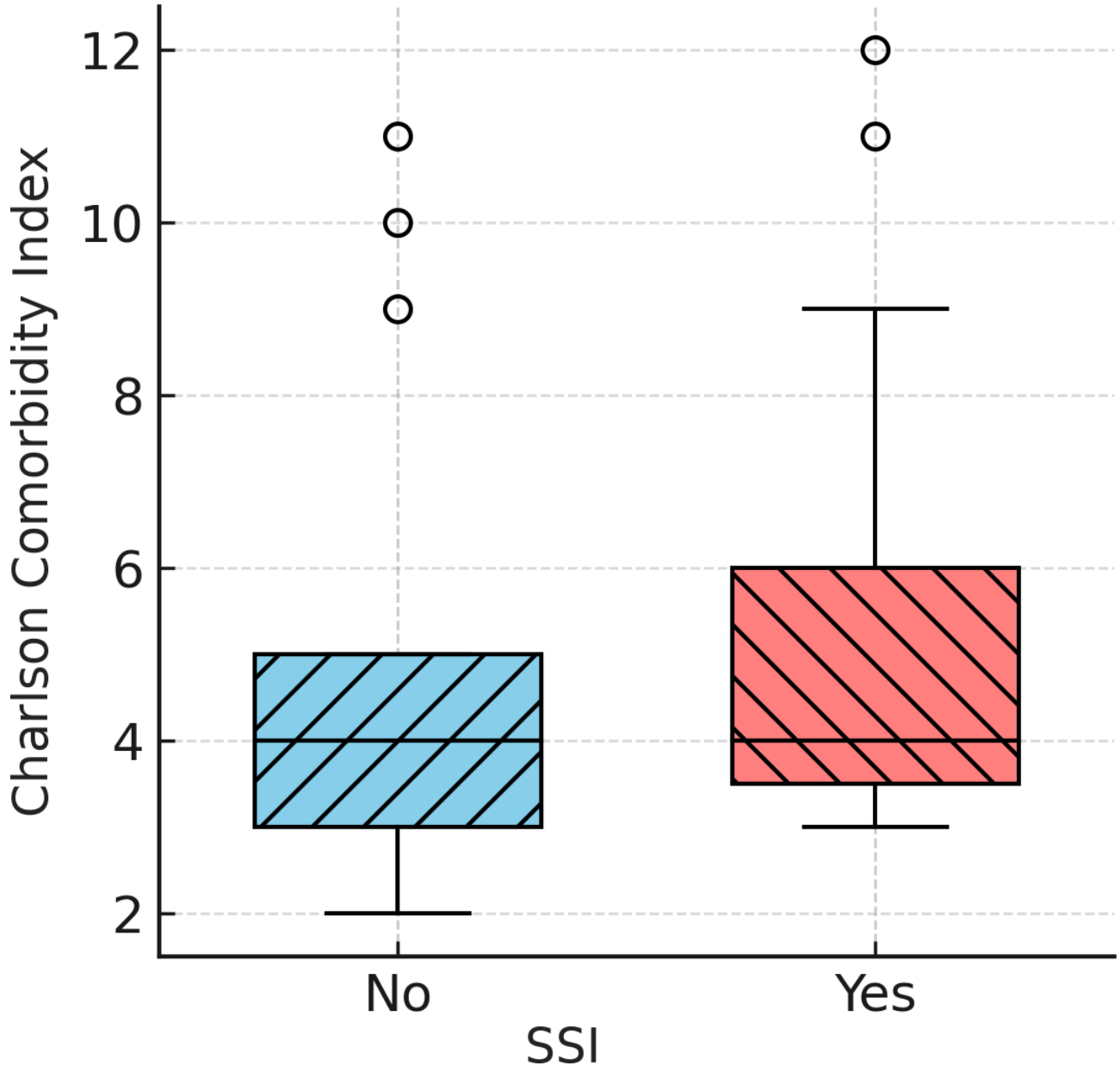
XSL•FO
RenderX

Figure 1. Distribution of postoperative infection according to New York Heart Association (NYHA) classification.



XSL•FO
RenderX

Figure 2. Comparison of the Charlson Comorbidity Index between patients with and without postoperative infection.



XSL•FO
RenderX

Table . Intraoperative variables in patients with and without postoperative infection.

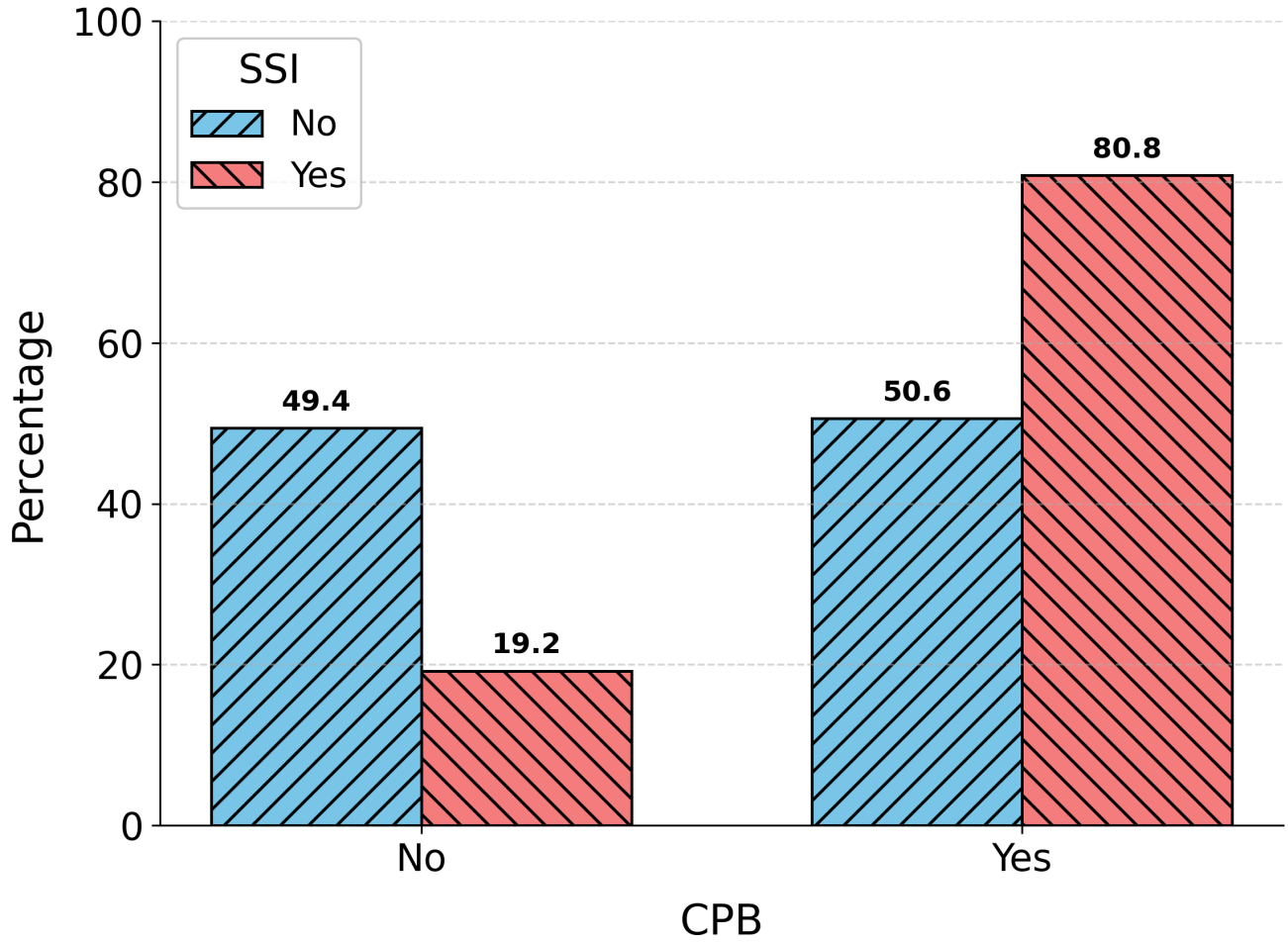
Variables	Infection		*P* value
	No (n=441)	Yes (n=26)	
Number of mammary arteries, n (%)			.52[a]
0	10 (2.3)	0 (0)	
1	298 (67.6)	20 (76.9)	
2	133 (30.2)	6 (23.1)	
Cardiopulmonary bypass, n (%)	223 (50.6)	21 (80.8)	.003[a]
Cardiopulmonary bypass time (min)[b], median (IQR)	85 (65-105)	85 (71-128)	.02[c]
Total surgery time (min), median (IQR)	270 (217-325)	303 (250-359)	<.001[c]

[a]Pearson chi-square test.

[b]Cardiopulmonary bypass duration was evaluated only among patients who underwent cardiopulmonary bypass (223 patients without surgical site infection and 21 patients with surgical site infection).

[c]Mann-Whitney *U* test.

Figure 3. Distribution of postoperative infection according to the use of cardiopulmonary bypass (CPB) during the procedure.



XSL•FO
RenderX

Table . Results for the postoperative variables in patients with and without postoperative infection.

Variables	Infection		*P* value
	No (n=441)	Yes (n=26)	
Bleeding, n (%)	4 (0.9)	2 (7.7)	.04[a]
Mechanical ventilation duration (h), median (IQR)	8 (5-14)	18 (8-21)	.001[b]
Length of stay (d), median (IQR)	7 (6-11)	41 (23-55)	<.001[b]
Length of stay in the ICU[c] (d), median (IQR)	3 (2-5)	5 (4-15)	<.001[b]
Reoperation during hospitalization, n (%)	13 (2.9)	13 (50.0)	<.001[a]

[a]Fisher exact test.

[b]Mann-Whitney *U* test.

[c]ICU: intensive care unit.

Table . Multiple logistic regression for factors independently associated with postoperative surgical site infection.

Variables	β coefficient[a] (SE)[b]	OR[c] (95% CI)	*P* value
Charlson Comorbidity Index	0.46 (0.12)	1.58 (1.25-1.99)	<.001
Cardiopulmonary bypass	1.59 (0.54)	4.89 (1.71-13.94)	.003

[a]Log-odds scale.

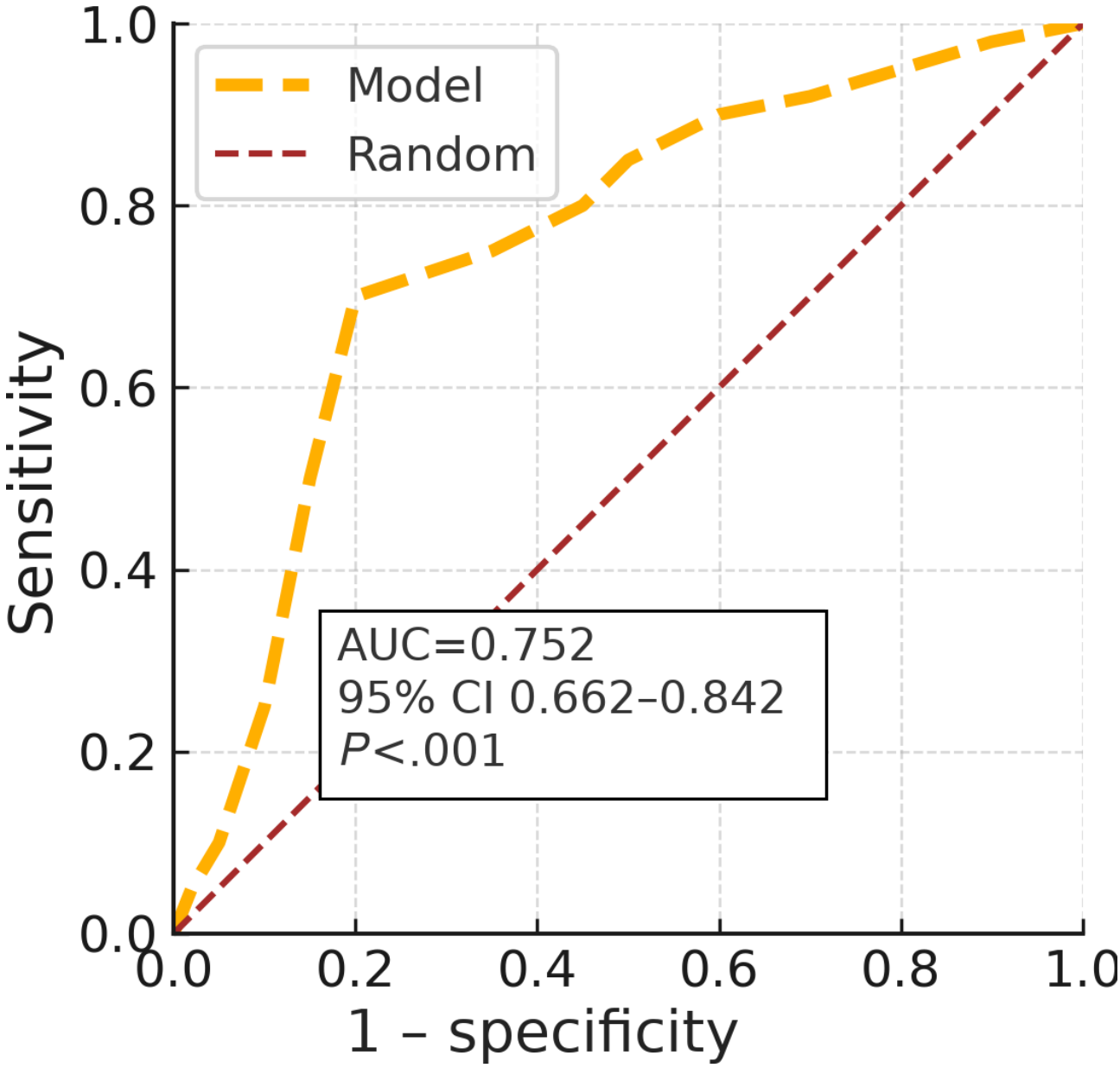
[b]Constant=–5.96 (SE 0.84).

[c]OR: odds ratio.

The CCI was associated with the outcome (OR 1.58, 95% CI 1.25 - 1.99; *P*<.001), indicating a 58% increase in the likelihood of infection with each additional point on the index. The use of CPB also proved to have a significant association (OR 4.89, 95% CI 1.71 - 13.94; *P*=.003). The model showed good discrimination, with an area under the receiver operating characteristic curve of 0.752 (95% CI 0.662 - 0.842; *P*<.001; Figure 4).

XSL•FO
RenderX

Figure 4. Receiver operating characteristic (ROC) curve of the predictive model for postoperative infection in high-risk surgical patients. AUC: area under the ROC curve.



Discussion

In this study, an SSI rate of 5.6% (26/467) was observed in patients with frailty undergoing myocardial revascularization. Although this was higher than the rates reported in previous studies [16,20-22], no study has addressed frailty in a population like the one we studied. Among the factors analyzed, the CCI and the use of CPB were identified as associated with SSI, especially CPB, which was independently associated with higher odds of SSI.

The association between frailty and adverse outcomes in surgery has been widely described. Studies have shown that patients with frailty have a higher risk of mortality, prolonged hospitalization, and infection complications [5,7,26]. In this study, even after adjusting for other variables, the CCI was significantly associated with the occurrence of SSI (OR 1.58, 95% CI 1.25 - 1.99; *P*<.001), which demonstrated that the accumulation of comorbidities associated with a decline in physiological reserves increases the association with SSI by 58% for each 1-point increase in the CCI. We additionally clarified how frailty was operationalized in the cohort and how its distribution may have limited discrimination within an already frail population.

The frailty score did not discriminate SSI risk within this cohort, likely because participants were selected for frailty and prefrailty by design, reducing variability in frailty burden and limiting discrimination for this end point. Rather than functioning as a within-cohort predictor, frailty in this study should be viewed as a biological substrate of vulnerability potentially linked to reduced immune responsiveness and diminished physiological reserve that may contribute to the susceptibility to SSI. Consequently, additional perioperative factors may better explain variation in SSI occurrence among patients already meeting frailty and prefrailty criteria.

XSL•FO
RenderX

The use of CPB also showed a significant association with infection, corroborating the literature that associates CPB with greater systemic inflammatory activation, increased tissue permeability, and transient immune suppression [27,28]. The same trend was observed for CPB time, which differed between the groups. The use of this technique and the intensity of the associated stress are important markers of operative complexity and greater exposure to infectious risks. Mechanistically, CPB is associated with complement activation and cytokine release, endothelial and glycocalyx injury, and transient immune dysfunction, which may impair wound healing and facilitate infection in vulnerable patients.

Reoperation was strongly associated with SSI in the unadjusted analysis. However, because reoperation was recorded after CABG and its temporal relationship with SSI diagnosis was not reported, it should be interpreted as a postoperative correlate rather than as a predictor of SSI. This finding corroborates those of studies showing that surgical revision, especially at an early stage, compromises anatomical barriers and favors colonization of deep tissue of the thorax [29]. Additionally, patients who require reoperation often have more severe clinical profiles and are subject to repeated stressors, including the use of invasive devices for prolonged periods.

Although variables such as age, BMI, and diabetes mellitus have traditionally been related to sternal infection [16], they did not prove to be significant in this study. This result appears to be associated with the fact that the analysis was specific to a population with frailty, in which factors associated with a drop in physiological reserve are more prominent.

Other relevant variables were NYHA functional class and length of mechanical ventilation, which were significantly associated with infection, similar to what has been described in the literature [17,30,31]. More advanced functional classes, such as NYHA class IV, showed a higher association with infection, possibly because they reflect greater previous hemodynamic compromise. This leads to longer intubation time in the postoperative period, which can hinder healing and predispose patients to infection.

The number of mammary artery grafts used did not influence the occurrence of infection; however, there are studies that demonstrate an increase in incidence when 2 mammary grafts are used [32,33]. The results of our study may encourage surgeons, when choosing grafts, to use a greater number of mammary grafts because of the benefit of durability and because the complications of SSI are similar whether one or more is used.

Length of hospital and ICU stay were associated with a higher occurrence of SSI, which indicates a negative cycle in which longer hospital stay increases the association with infection and the occurrence of infection increases the length of hospital stay. Longer hospital stays are associated with longer use of invasive devices, more contact with hospital pathogens, and changes in the patient's microbiota. In turn, the occurrence of infection contributes to an increase in hospital admissions, the use of antibiotics, the need for procedures secondary to infection, and delay in postoperative recovery. Consequently, guidelines associated with multiple actions have emerged with the aim to reduce the length of hospital stay and, subsequently, reduce the occurrence of SSI [34].

Makary et al [7] suggested incorporating frailty score into the set of risk predictors commonly used in clinical practice. Risk assessment in cardiovascular surgery is carried out using scores such as the EuroSCORE and the Society of Thoracic Surgeons score, which estimate risk without considering the decline in the individual's physiological reserves. The inclusion of the frailty score and prior assessment of comorbidities during the evaluation contribute to the early identification of patients who are more susceptible to complications, allowing for preventive actions through adjustments to the surgical technique, preoperative preparation of the patient's physical condition, and strict control of comorbidities.

This multicenter study used standardized criteria for assessing frailty and multivariable logistic regression, which showed that the CCI and CPB use were independently associated with SSI. Preoperative recognition of patients with frailty with a high number of comorbidities, especially those undergoing CPB, can guide prevention strategies and more intensive monitoring in the pre- and postoperative periods.

However, this study had some limitations. The number of SSI events was relatively small, which may have limited the stability of the multivariable model and increased the risk of overfitting. The retrospective use of medical records from multiple centers may have resulted in variation in the completeness and consistency of documentation for certain clinical variables, particularly SSI subtype, microbiological findings, and antibiotic treatment. Because of the limited number of SSI events, we did not adjust for center-level effects, and residual confounding due to interinstitutional variation in perioperative care and SSI surveillance cannot be excluded. The composite SSI end point included events with different clinical severity (superficial infection, deep infection, mediastinitis, and dehiscence). Because of the limited number of events in each subtype, we were unable to perform robust subtype-specific modeling; thus, associations may differ across SSI phenotypes. We assessed discrimination (area under the receiver operating characteristic curve) but did not formally evaluate model calibration; therefore, the clinical applicability of predicted risks should be interpreted cautiously. Although CPB assignment was randomized in the parent trial, the present analysis included only centers with complete eligible data; therefore, postrandomization exclusions and center selection may have introduced selection bias and limited generalizability. We did not formally quantify multicollinearity using the variance inflation factor or correlation coefficients; therefore, residual collinearity among operative variables cannot be excluded.

We conclude that factors such as NYHA functional class IV, CCI, CPB time, bleeding, duration of mechanical ventilation, need for reoperation, use of CPB, and length of hospital and ICU stay showed an unadjusted association with SSI. Additionally, we demonstrated that the CCI and the use of CPB were independently associated with SSI in the multivariable model within 30 days after CABG.

XSL•FO
RenderX

Acknowledgments

All collaborators who were part of the FRAGILE project, the original study, were responsible for providing assistance to patients, maintained the database, and contributed to this study.

The authors used ChatGPT (OpenAI) under full human supervision for administrative correspondence and language and formatting support during copyediting. Generative AI was not used in the study design, data collection, statistical analysis, or interpretation of the results. The authors reviewed and approved all changes and take full responsibility for the final manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MG, FJ
Data curation: MG, OM, BM, DW
Investigation: MG, OM, NR, WB, MB, DW, LB, AH
Software: BM
Supervision: FJ
Writing—original draft: MG
Writing—review and editing: FJ

Conflicts of Interest

None declared.

References

1. DataBank | World development indicators. The World Bank Group. URL: https://databank.worldbank.org/reports.aspx?source=2&series=SP.DYN.LE00.MA.IN&country= [accessed 2025-01-11]
2. The top 10 causes of death. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2025-03-26]
3. Feitosa-Filho GS, Peixoto JM, Pinheiro JES, et al. Updated geriatric cardiology guidelines of the Brazilian Society of Cardiology - 2019. Arq Bras Cardiol 2019 Jun 6;112(5):649-705. [doi: 10.5935/abc.20190086] [Medline: 31188969]
4. Kuller LH, Lopez OL, Mackey RH, et al. Subclinical cardiovascular disease and death, dementia, and coronary heart disease in patients 80+ years. J Am Coll Cardiol 2016 Mar 8;67(9):1013-1022. [doi: 10.1016/j.jacc.2015.12.034] [Medline: 26940919]
5. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
6. Bandeen-Roche K, Xue QL, Ferrucci L, et al. Phenotype of frailty: characterization in the women's health and aging studies. J Gerontol A Biol Sci Med Sci 2006 Mar;61(3):262-266. [doi: 10.1093/gerona/61.3.262] [Medline: 16567375]
7. Makary MA, Segev DL, Pronovost PJ, et al. Frailty as a predictor of surgical outcomes in older patients. J Am Coll Surg 2010 Jun;210(6):901-908. [doi: 10.1016/j.jamcollsurg.2010.01.028] [Medline: 20510798]
8. Polanczyk CA, Marcantonio E, Goldman L, et al. Impact of age on perioperative complications and length of stay in patients undergoing noncardiac surgery. Ann Intern Med 2001 Apr 17;134(8):637-643. [doi: 10.7326/0003-4819-134-8-200104170-00008] [Medline: 11304103]
9. Martin SS, Aday AW, Almarzooq ZI, et al. 2024 Heart disease and stroke statistics: a report of US and global data from the American Heart Association. Circulation 2024 Feb 20;149(8):e347-e913. [doi: 10.1161/CIR.0000000000001209] [Medline: 38264914]
10. Informações de saúde (TABNET). DATASUS. URL: https://datasus.saude.gov.br/informacoes-de-saude-tabnet/ [accessed 2024-10-23]
11. Bowdish ME, D'Agostino RS, Thourani VH, et al. STS Adult Cardiac Surgery Database: 2021 update on outcomes, quality, and research. Ann Thorac Surg 2021 Jun;111(6):1770-1780. [doi: 10.1016/j.athoracsur.2021.03.043] [Medline: 33794156]
12. Kim KM, Arghami A, Habib R, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2022 update on outcomes and research. Ann Thorac Surg 2023 Mar;115(3):566-574. [doi: 10.1016/j.athoracsur.2022.12.033] [Medline: 36623634]

XSL•FO
RenderX

13. Lazar HL, Fitzgerald CA, Ahmad T, et al. Early discharge after coronary artery bypass graft surgery: are patients really going home earlier? J Thorac Cardiovasc Surg 2001 May;121(5):943-950. [doi: 10.1067/mtc.2001.113751] [Medline: 11326238]
14. Taggart DP, Altman DG, Gray AM, et al. Randomized trial of bilateral versus single internal-thoracic-artery grafts. N Engl J Med 2016 Dec 29;375(26):2540-2549. [doi: 10.1056/NEJMoa1610021] [Medline: 27959712]
15. Gummert JF, Barten MJ, Hans C, et al. Mediastinitis and cardiac surgery--an updated risk factor analysis in 10,373 consecutive adult patients. Thorac Cardiovasc Surg 2002 Apr;50(2):87-91. [doi: 10.1055/s-2002-26691] [Medline: 11981708]
16. Chen D, Zhang J, Wang Y, et al. Risk factors for sternal wound infection after open-heart operations: a systematic review and meta-analysis. Int Wound J 2024 Mar;21(3):e14457. [doi: 10.1111/iwj.14457] [Medline: 37909266]
17. Loop FD, Lytle BW, Cosgrove DM, et al. J. Maxwell Chamberlain memorial paper. Sternal wound complications after isolated coronary artery bypass grafting: early and late mortality, morbidity, and cost of care. Ann Thorac Surg 1990 Feb;49(2):179-186. [doi: 10.1016/0003-4975(90)90136-t] [Medline: 2306138]
18. Losanoff JE, Richman BW, Jones JW. Disruption and infection of median sternotomy: a comprehensive review. Eur J Cardiothorac Surg 2002 May;21(5):831-839. [doi: 10.1016/s1010-7940(02)00124-0] [Medline: 12062271]
19. Lemaignen A, Birgand G, Ghodhbane W, et al. Sternal wound infection after cardiac surgery: incidence and risk factors according to clinical presentation. Clin Microbiol Infect 2015 Jul;21(7):674.e11. [doi: 10.1016/j.cmi.2015.03.025] [Medline: 25882356]
20. Montague NT 3rd, Kouchoukos NT, Wilson TA, et al. Morbidity and mortality of coronary bypass grafting in patients 70 years of age and older. Ann Thorac Surg 1985 Jun;39(6):552-557. [doi: 10.1016/s0003-4975(10)61997-2] [Medline: 3873921]
21. Piątek J, Kędziora A, Konstanty-Kalandyk J, et al. Risk factors for in-hospital mortality after coronary artery bypass grafting in patients 80 years old or older: a retrospective case-series study. PeerJ 2016;4:e2667. [doi: 10.7717/peerj.2667] [Medline: 27920951]
22. Sun L, Zhou M, Ji Y, Wang X, Wang X. Off-pump versus on-pump coronary artery bypass grafting for octogenarians: a meta-analysis involving 146 372 patients. Clin Cardiol 2022 Apr;45(4):331-341. [doi: 10.1002/clc.23794] [Medline: 35266173]
23. Mejía OA, Sá MP, Deininger MO, et al. Off-pump versus on-pump coronary artery bypass grafting in frail patients: study protocol for the FRAGILE multicenter randomized controlled trial. Braz J Cardiovasc Surg 2017;32(5):428-434. [doi: 10.21470/1678-9741-2017-0196] [Medline: 29211225]
24. Performance measure descriptions. The Society of Thoracic Surgeons. URL: https://www.sts.org/quality-safety/performance-measures/descriptions [accessed 2026-09-03]
25. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
26. Sepehri A, Beggs T, Hassan A, et al. The impact of frailty on outcomes after cardiac surgery: a systematic review. J Thorac Cardiovasc Surg 2014 Dec;148(6):3110-3117. [doi: 10.1016/j.jtcvs.2014.07.087] [Medline: 25199821]
27. Navia D, Espinoza J, Vrancic M, et al. Bilateral internal thoracic artery grafting in elderly patients: any benefit in survival? J Thorac Cardiovasc Surg 2022 Aug;164(2):542-549. [doi: 10.1016/j.jtcvs.2020.09.101] [Medline: 33127090]
28. Sedrakyan A, Wu AW, Parashar A, Bass EB, Treasure T. Off-pump surgery is associated with reduced occurrence of stroke and other morbidity as compared with traditional coronary artery bypass grafting: a meta-analysis of systematically reviewed trials. Stroke 2006 Nov;37(11):2759-2769. [doi: 10.1161/01.STR.0000245081.52877.f2] [Medline: 17008617]
29. Schimmer C, Reents W, Berneder S, et al. Prevention of sternal dehiscence and infection in high-risk patients: a prospective randomized multicenter trial. Ann Thorac Surg 2008 Dec;86(6):1897-1904. [doi: 10.1016/j.athoracsur.2008.08.071] [Medline: 19022005]
30. Ridderstolpe L, Gill H, Granfeldt H, Ahlfeldt H, Rutberg H. Superficial and deep sternal wound complications: incidence, risk factors and mortality. Eur J Cardiothorac Surg 2001 Dec;20(6):1168-1175. [doi: 10.1016/s1010-7940(01)00991-5] [Medline: 11717023]
31. Lu JC, Grayson AD, Jha P, Srinivasan AK, Fabri BM. Risk factors for sternal wound infection and mid-term survival following coronary artery bypass surgery. Eur J Cardiothorac Surg 2003 Jun;23(6):943-949. [doi: 10.1016/s1010-7940(03)00137-4] [Medline: 12829070]
32. Parisian Mediastinitis Study Group. Risk factors for deep sternal wound infection after sternotomy: a prospective, multicenter study. J Thorac Cardiovasc Surg 1996 Jun;111(6):1200-1207. [doi: 10.1016/S0022-5223(96)70222-2] [Medline: 8642821]
33. Zacharias A, Habib RH. Factors predisposing to median sternotomy complications. Deep vs superficial infection. Chest 1996 Nov;110(5):1173-1178. [doi: 10.1378/chest.110.5.1173] [Medline: 8915216]
34. Engelman DT, Ben Ali W, Williams JB, et al. Guidelines for perioperative care in cardiac surgery: Enhanced Recovery After Surgery Society recommendations. JAMA Surg 2019 Aug 1;154(8):755-766. [doi: 10.1001/jamasurg.2019.1153] [Medline: 31054241]

XSL•FO
RenderX

Abbreviations

CABG: coronary artery bypass grafting
CCI: Charlson Comorbidity Index
CPB: cardiopulmonary bypass
FRAGILE: off-pump versus on-pump coronary artery bypass grafting in frail patients
ICU: intensive care unit
NYHA: New York Heart Association
OR: odds ratio
SSI: surgical site infection

Edited by N Ainani; submitted 20.Jun.2025; peer-reviewed by DA Hanafy, MK Toker; revised version received 23.Jun.2026; accepted 12.Aug.2026; published 16.Oct.2026.

Please cite as:
Guerrieri M, Mejia O, Meneghini B, Rodrigues N, Burato W, Barros M, Wernsbach D, Bartelega L, Hueb A, Jatene F
Influence of Perioperative Variables on Surgical Site Infection in Patients With Frailty and Prefrailty Undergoing Coronary Artery Bypass Grafting: Retrospective Multicenter Study
JMIR Cardio 2026;10:e79408
URL: https://cardio.jmir.org/2026/1/e79408
doi:10.2196/79408

© Mauricio Guerrieri, Omar Mejia, Bianca Meneghini, Natali Rodrigues, William Burato, Maisa Barros, Daniel Wernsbach, Lucas Bartelega, Alexandre Hueb, Fabio Jatene. Originally published in JMIR Cardio (https://cardio.jmir.org), 16.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Literacy in Elective Open-Heart Surgery Patients: Cross-Sectional Study

Rikke Daugaard[1,2], MSHS; Thomas Maribo[3,4], PhD; Britt Borregaard[5,6,7], PhD; Ditte Sommerlund Skydt[8], MSHS; Kristina Hindhede Bech[5], MHPT; Kjersti Alexandra Skovli[9,10], Dr med; Ivy Susanne Modrau[1,2], PhD

[1]Department of Clinical Medicine, Faculty of Health, Aarhus University, Vennelyst Boulevard 4, Aarhus, Denmark

[10]Department of Clinical Medicine, Faculty of Medicine, University of Aalborg, Aalborg, Denmark

[2]Department of Cardiothoracic and Vascular Surgery, Aarhus University Hospital, Aarhus North, Denmark

[3]Department of Public Health, Faculty of Health, Aarhus University, Aarhus, Denmark

[4]DEFACTUM, Corporate Quality, Central Denmark Region, Aarhus, Denmark

[5]Department of Cardiothoracic and Vascular Surgery, Odense University Hospital, Odense, Denmark

[6]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[7]Department of Cardiology, Odense University Hospital, Odense, Denmark

[8]Esbjerg Municipal Health Centres, Esbjerg Municipality, Esbjerg, Denmark

[9]Department of Cardiothoracic Surgery, Aalborg University Hospital, Aalborg, Denmark

Corresponding Author:
Rikke Daugaard, MSHS
Department of Clinical Medicine, Faculty of Health, Aarhus University, Vennelyst Boulevard 4, Aarhus, Denmark

Abstract

Background: Digital health solutions play a key role in health care, but their safe and effective use depends on patients' digital health literacy. While digital health solutions are beneficial for patients with cardiac disease, disparities in digital health literacy may limit access, particularly for patients undergoing cardiac surgery with complex care and psychological challenges. Unaddressed, these disparities could exacerbate inequalities in accessing beneficial digital services. Denmark's advanced digital health care system provides a unique context to evaluate digital health literacy.

Objective: This study aimed to assess digital health literacy levels in patients scheduled for elective open-heart surgery and examine associations with sociodemographic factors and concurrent health issues.

Methods: We conducted a cross-sectional survey of consecutive patients scheduled for elective open-heart surgery at 3 university hospitals covering approximately two-thirds of Denmark's population. Patients with cognitive impairment or language barriers preventing completion of the questionnaire were excluded. The questionnaire was administered in paper form by medical staff during preoperative consultations. Digital health literacy was assessed using the validated 8-item eHealth Literacy Scale (eHEALS; range 8 - 40), along with 2 additional questions from the validated Danish version assessing the perceived importance and usefulness of online health information. Sociodemographic data collected included age, gender, cohabitation status, social support for technology use, educational level, and number of concurrent health issues. Descriptive and comparative analyses examined associations between eHEALS scores and sociodemographic variables and health issues. Exploratory subscale analyses evaluated the 3 eHEALS domains—awareness, skills, and ability to evaluate online health information—to identify areas in which patients may require additional support.

Results: Of 576 eligible patients, 313 (54.3%) completed the survey between February 2024 and July 2024. Response rates varied across sites: 71.1% (133/187), 58.5% (134/229), and 28.8% (46/160) in sites 1, 2, and 3, respectively. Nonresponse was primarily due to logistical challenges during preoperative consultations, with only a few patients excluded because of cognitive impairment, language barriers, or refusal. The median eHEALS score was 30 (IQR 27 - 32), indicating generally high digital health literacy scores (cutoff score ≥26). Scores were negatively correlated with age (Spearman ρ=–0.18; P=.002) and positively associated with educational level (Kruskal-Wallis test: χ^2_2=17.0; P<.001). No substantial associations were observed for gender, cohabitation status, social support for technology use, or number of concurrent health issues. Exploratory subscale analyses suggested that patients felt least confident in evaluating the quality and relevance of online information, highlighting a potential focus for tailored support.

Conclusions: Patients scheduled for elective open-heart surgery generally reported high digital health literacy levels, but challenges remain in critically appraising digital health information. Younger age and higher educational levels were associated with higher self-reported digital health literacy, but the association was modest. This underscores the need for individual assessment to identify patients who may benefit from tailored support.

XSL•FO
RenderX

(JMIR Cardio 2026;10:e83454) doi:10.2196/83454

KEYWORDS

cardiovascular diseases; thoracic surgery; elective surgical procedures; health literacy; telemedicine; computer literacy; health communication; sociodemographic factors; surveys; questionnaires

Introduction

Digital health has become an essential component of modern health care delivery. Over the past decade, digital health solutions and telemedicine have steadily evolved, transforming the way in which health care is accessed and delivered. The COVID-19 pandemic significantly accelerated this development as physical consultations were limited and the need for remote solutions intensified [1,2]. Today, digital health forms an integral part of routine health care across a range of clinical settings.

These solutions are particularly valuable in outpatient cardiac rehabilitation. They can help overcome time and travel barriers during structured programs (phase 2) and support long-term recovery and lifestyle maintenance (phase 3) [3]. Digital health literacy, defined as the ability to seek, find, understand, and appraise digital health information [4], is essential when implementing digital health solutions as low digital health literacy is associated with reduced awareness and use of digital services [5]. Moreover, addressing patients' digital health literacy may be important to ensure more inclusive implementation of these solutions, enabling broader engagement and reducing inequities [6]. Digital health interventions have been associated with improvements in health-related quality of life, as measured using validated instruments such as the HeartQoL questionnaire in cardiac populations, and with increased physical activity in patients with chronic conditions [7,8]. A recent systematic review in community-dwelling older adults reported relatively low digital health literacy as measured using the eHealth Literacy Scale (eHEALS) instrument [9]. Lower socioeconomic status and self-reported lack of digital skills have also been reported to affect digital health literacy negatively [10]. This may contribute to health disparities if health care systems do not account for patients' digital health literacy when designing and implementing digital solutions. Consequently, patients with lower digital health literacy may face unequal opportunities to enhance their health and well-being [11]. Therefore, it is important to investigate whether sociodemographic factors are associated with digital health literacy in patients with cardiac disease, as identifying such associations could inform targeted strategies to support effective engagement with digital health interventions.

In this study, we aimed to assess digital health literacy levels in patients scheduled for open-heart surgery and examine differences across sociodemographic factors and concurrent health issues.

Methods

Design, Study Setting, Population, and Data Collection

This cross-sectional survey collected data via a questionnaire administered at Aarhus University Hospital, Aalborg University Hospital, and Odense University Hospital, covering 3 regions that together represent approximately two-thirds of the Danish population. Adult patients scheduled for elective open-heart surgery between February 8, 2024, and July 5, 2024, were invited to participate. Nurses or medical secretaries distributed the questionnaire in paper form to eligible patients during their preoperative consultation and collected it anonymously the day before surgery. Inclusion criteria were patients aged ≥18 years scheduled for elective open-heart surgery. Patients unable to read or understand Danish or deemed by staff to be cognitively unable to complete the questionnaire were excluded. The responsible staff members assessed whether patients were suitable to participate.

The reporting of this study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [12].

Questionnaire Measures: Digital Health Literacy, Sociodemographics, and Data on Health Issues

The questionnaire consisted of 2 sections: one assessing patients' self-reported digital health literacy and the other gathering their sociodemographic information and data on health issues.

We used the eHEALS to assess self-reported digital health literacy [13]. The eHEALS has been validated in multiple languages, including Danish, and applied across various demographic groups [14]. The eHEALS consists of 8 items that assess the patient's self-perceived ability to access, evaluate, and apply health information found on the internet [15]. Respondents rate each item on a 5-point Likert scale ranging from "strongly disagree" (1 point) to "strongly agree" (5 points). Scores for each item are summed, yielding a total score of 8 to 40 points. A higher score indicates a higher level of digital health literacy, with ≥26 considered the cutoff between high and low digital health literacy levels [16-18]. Two supplementary items included in the validated Danish eHEALS assessed patients' perceived importance and usefulness of digital health (not included in the total score) to provide contextual information on attitudes toward digital health. Furthermore, we analyzed the digital health literacy assessment according to the 3-factor model as defined by Sudbury-Riley et al [16]: "awareness of internet health resources" (items 1 and 2), "skills and behavior needed to access health information" (items 3-5), and "self-belief in ability to evaluate health resources" (items 6-8). This domain-specific approach allows for a more nuanced evaluation of patients' competencies in accessing, using, and appraising online health information rather than relying solely on a total eHEALS score [16].

The second section of the questionnaire included questions on self-reported sociodemographic factors and health issues and was developed based on known risk factors associated with low digital health literacy in patients with cardiac disease [17-19]. Sociodemographic factors included age, gender, educational level, cohabitation status, and social support for technology use.

XSL•FO
RenderX

Patients were also asked to specify any health issues beyond the cardiac disease. Educational levels were categorized according to the International Standard Classification of Education as high school (0-2 years of post–high school education), undergraduate (3-4 years of post–high school education), and postgraduate (>4 years of post–high school education) [20]. Cohabitation status was determined as either being a cohabitant or living alone. Social support for technology use was defined as the availability of help from friends or family in addressing technological challenges when necessary. Health issues beyond the primary cardiac diagnosis were represented by a checklist of the most common health issues, including diabetes, pulmonary disease, vascular disease, back and muscle pain, asthma, rheumatoid arthritis, osteoporosis, migraine or severe headache, and chronic kidney disease [21]. Patients could specify any additional diseases in a free-text field. Health issues were categorized as "none," "one," or "two or more."

Data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University), hosted by Aarhus University [22].

Statistical Methods

Patients with missing items on the eHEALS were excluded to prevent underestimation of the score. On the basis of prior studies, a nonnormal distribution of eHEALS scores was anticipated, and therefore, nonparametric tests were expected to be appropriate. Normality of the data was visually assessed before selecting the statistical approach. Log transformation was attempted but did not normalize the distribution; thus, nonparametric tests were applied.

We assessed differences in eHEALS scores using the Mann-Whitney *U* test for dichotomous variables and the Kruskal-Wallis test for categorical variables with 3 levels. The correlation between age and eHEALS scores was evaluated using the Spearman rank correlation, with values interpreted as weak (0.10 - 0.30), moderate (0.31 - 0.50), or strong (>0.50), consistent with commonly accepted conventions in the literature [23]. The 3 subscales of the digital health literacy assessment were analyzed by calculating medians and IQRs for each subscale.

All analyses were performed using the R statistical software (version 4.1.2; R Foundation for Statistical Computing) and RStudio (version 2024.12.0; Posit PBC).

Ethical Considerations

The questionnaire included a preamble outlining the study's objectives, the voluntary nature of participation, and the confidentiality and anonymity of respondents. Participants were informed that no personally identifiable data (eg, name, civil registration number, or contact information) were collected. All data were collected and stored in a secure environment accessible only to members of the research team and handled in accordance with the General Data Protection Regulation and institutional data protection policies. Data were analyzed and reported in aggregate form only, ensuring that individual participants could not be identified. Informed consent was implied through completion and submission of the questionnaire. Participants were informed that they could discontinue participation at any time before submission without any consequences and that, due to the anonymous nature of the survey, withdrawal after submission would not be possible. No financial or other compensation was provided for participation. This study was conducted in accordance with the Declaration of Helsinki. The Research Ethics Committee of the Danish Regions waived the need for study approval as the study was exempt from registration under Danish law (Danish Committee Act on Research Ethics Review of Health Research Projects §14 subsection 2; case number: 2400348) [24].

Results

General Characteristics

In total, 54.3% (313/576) of the eligible respondents were included. Separate response rates among eligible patients at the 3 sites were 71.1% (133/187), 58.5% (134/229), and 28.8% (46/160). Baseline characteristics of the study population are summarized in Table 1.

XSL•FO
RenderX

Table . Sociodemographic characteristics and eHealth Literacy Scale (eHEALS) score of study participants (n=313).

Participant characteristic	Values
Sociodemographics	
Age (y), mean (SD)	65.8 (11)
Gender (missing: 2; n=311), n (%)	
Male	232 (74.6)
Female	79 (25.4)
Educational level (missing: 8; n=305), n (%)	
High school	92 (30.2)
Undergraduate	170 (55.7)
Postgraduate	43 (14.1)
Cohabitation status (missing: 0), n (%)	
Living alone	78 (24.9)
Cohabiting	235 (75.1)
Close relationships (missing: 3; n=310), n (%)	
Reported close personal relationships	276 (89.0)
No close personal relationships	34 (11.0)
Concurrent health issues (missing: 16; n=297), n (%)	
Number of health issues	
None	135 (45.5)
1	102 (34.3)
≥2	60 (20.2)
Digital health literacy level, median (IQR)	
eHEALS score (8-40)	30 (27-32)

As anticipated, eHEALS data were not normally distributed (Figure 1). The median eHEALS score was 30 (IQR 27 - 32), and 91.8% (267/291) of the participants scored 26 or higher, indicating that participants generally possessed high self-reported digital health literacy levels. A total of 7% (22/313) of the questionnaires were discarded due to missing items. As a complement to the eHEALS, patients were asked about their opinions on the importance and usefulness of digital health. Just over half of the participants rated the internet as “useful” for health-related decisions (160/291, 55%) and considered access to online health information to be “important” (157/291, 54%) to them (Figure 2).

XSL•FO
RenderX

Figure 1. Distribution of total eHealth Literacy Scale (eHEALS) scores among participants (n=291).

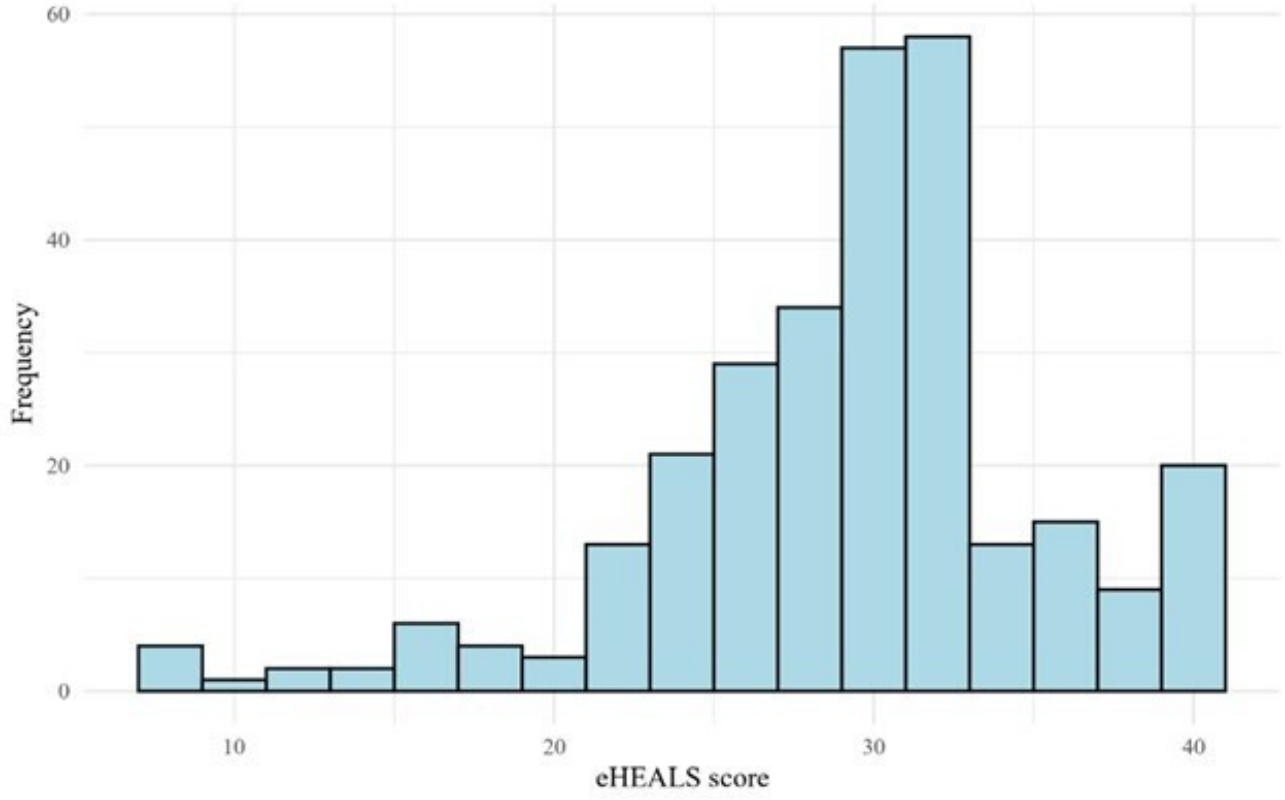
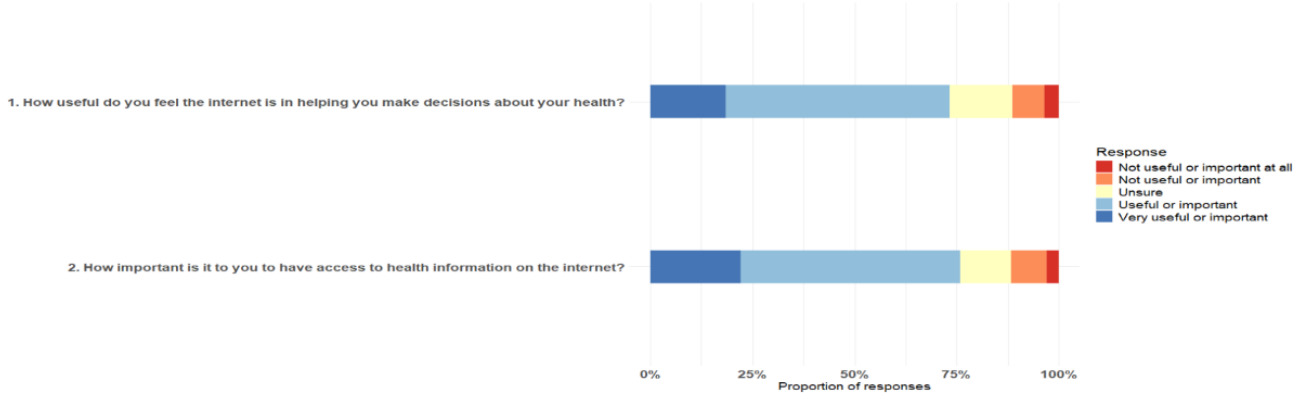


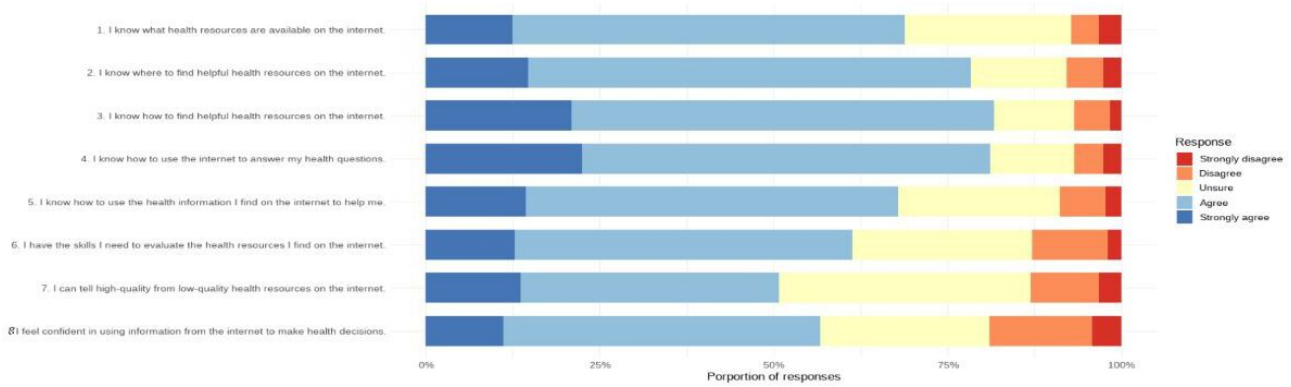
Figure 2. Distribution of responses to the 2 questions supplementing the eHealth Literacy Scale on the usefulness and importance of online health information.



The 3 subscales of the digital health literacy assessment were analyzed separately. The median scores were 8 (IQR 8-8) for "awareness" (2 items; maximum score=10), 12 (IQR 10-13) for "skills" (3 items; maximum score=15), and 12 (IQR 10-12) for "evaluate" (3 items; maximum score=15). These results indicate similar levels across all 3 domains. The proportion of responses across the 5-point Likert scale for each question is illustrated in Figure 3. Across all 8 eHEALS items, most patients selected "agree," indicating an overall high perceived digital health literacy. Response distributions across items suggest lower confidence in the "evaluate" domain. Notably, for item 7 ("I can tell high-quality from low-quality health resources on the Internet"), a substantial proportion of participants responded with "unsure," indicating limited confidence in critically appraising online health information.

XSL•FO
RenderX

Figure 3. Distribution of eHealth Literacy Scale responses grouped according to the 3-factor model: “awareness” (items 1-2), “skills” (items 3-5), and “evaluate” (items 6-8).



Sociodemographic Factors, Health Issues, and eHEALS Scores

Most participants (276/310, 89.0%) reported having close relationships that could assist them with technical issues. The most common sources of social support were children and spouses.

Age was significantly negatively correlated with eHEALS scores (Spearman ρ=–0.18; *P*=.002), reflecting a weak decline in self-reported digital health literacy as age increases. Additionally, a significant difference in eHEALS scores was observed across educational levels, with higher educational level corresponding to higher eHEALS scores (Kruskal-Wallis test: χ^2_2=17.0; *P*<.001).

No statistically significant differences in eHEALS scores were found across gender, cohabitation status, presence of close relationships, or number of health issues. The results are presented in Table 2 and as box plots and scatterplots in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Impact of sociodemographic factors and health issues on self-reported digital health literacy level.

Variables	eHEALS[a] score (8-40), median (IQR)	Test statistic	*P* value
Age (y)	30 (27-32)	–0.18[b]	.002
Gender		7426.5[c]	.46
Male	30 (26-32)		
Female	30 (29-32)		
Educational level		17.0[d]	<.001
High school	29 (25-32)		
Undergraduate	31 (27-32)		
Postgraduate	33 (30-39)		
Cohabitation status		8317.5[c]	.56
Living alone	30 (27-32)		
Cohabiting	31 (27-32)		
Close relationships		3992.5[c]	.47
Yes	30 (27-32)		
No	30 (23-32)		
Health issues		0.1[d]	.94
None	31 (27-32)		
1	30 (27-32)		
≥2	30 (26-34)		

[a]eHEALS: eHealth Literacy Scale.

[b]Spearman rank correlation.

[c]Mann-Whitney *U* test (*W* statistic).

[d]Kruskal-Wallis test (chi-square value with *df* of 2).

Discussion

Principal Findings

In this cross-sectional study of patients referred to elective open-heart surgery, we found that self-reported digital health literacy was generally high, with younger age and higher educational level associated with higher eHEALS scores. No significant differences were observed for gender, cohabitation status, presence of close relationships, or number of concurrent health issues. Despite high overall eHEALS scores, many patients reported limited confidence in their ability to critically appraise online health information. To our knowledge, this is the first assessment of self-reported digital health literacy in a preoperative cardiac surgery population. Evaluating digital competencies is particularly important in this high-risk group, which faces existential threats and complex care demands. Here, digital health literacy is essential to support informed decision-making, perioperative care, and postoperative cardiac rehabilitation.

Our finding related to younger age and higher self-reported digital health literacy is consistent with prior evidence from a systematic review of 17 studies in the general population [25], as well as a cross-sectional study of individuals with moderate to high cardiovascular risk [17]. While some studies have not found an association between age and self-reported digital health literacy measured using the eHEALS, their generalizability is limited due to factors such as small, selective samples, as is the case in 2 Danish studies on patients with cardiac disease [19,26]; selection bias from motivated participants with cardiac disease in a randomized controlled trial [27]; and limited applicability to the European context of an Asian cohort from the general population [28]. Recent evidence shows a general trend toward narrowing age-related gaps in digital health literacy [9,10,18], likely due to increased digitalization and technology adoption among older adults. Our findings indicate that age-related disparities in digital health literacy persist even in Denmark, a highly digitalized country where approximately 90% of older adults use the internet and own smartphones [29]. This potential among older adults underscores the ongoing need to provide adequate support to this group to ensure that they benefit from digital health solutions [27,30].

Consistent with previous research [17,25], patients with higher educational levels reported high digital health literacy levels measured using the eHEALS. However, patients with lower educational levels also demonstrated moderately high scores, indicating less disparity in digital health competencies within this population than anticipated.

XSL•FO
RenderX

No association was found between the number of health issues and eHEALS scores. Although direct evidence is lacking, it is plausible that repeated engagement with digital health services in a highly digitalized system may help maintain or enhance digital health literacy among patients with multiple chronic conditions, potentially counteracting any negative effects of higher disease burden [31].

The findings highlight that, although certain sociodemographic factors may influence digital competencies, they do not fully explain individual variation in digital health literacy. In a health care context in which digital solutions are increasingly integrated into perioperative and rehabilitative care, awareness of these individual differences remains essential. Tailoring digital health solutions to patients' varying levels of digital competence can enhance accessibility, promote equity, and help prevent the deepening of existing health disparities [32].

Assessing digital health literacy in cardiac surgery patients is important when identifying those who may require support to effectively engage with digital health information during the preoperative and rehabilitation phases. The eHEALS was selected for its validated, concise format, which allows for practical screening in the stressful, time-constrained preoperative setting. Although it provides only a focused assessment of patients' ability to seek, find, and appraise online health information, it remains suitable for identifying patients who might benefit from additional guidance. More comprehensive instruments such as the Digital Health Literacy Instrument [33] or the eHealth Literacy Questionnaire [34] offer broader assessments but are less practical for routine clinical use.

The widely used eHEALS provides a brief assessment of perceived digital health literacy, but as it was developed in 2006, it does not reflect the competencies needed to engage with modern digital health technologies. Specifically, it does not address users' ability to interpret and use data from mobile health apps and wearables, evaluate the credibility of health information on social media and video-sharing platforms, or navigate secure patient portals to communicate with health care providers. Furthermore, it lacks items related to privacy, data security, and critical appraisal of digital sources. Therefore, high eHEALS scores may primarily reflect proficiency in health 1.0 skills (eg, searching for health information online) while masking potential gaps in the more interactive and participatory health 2.0 skills required in contemporary digital health ecosystems [35]. This highlights the need for updated instruments that capture these multidimensional competencies and remain practical for use in clinical research and routine care. Screening alone cannot address digital health disparities; effective solutions also demand enhanced digital health literacy among health care providers and system-level support [36]. However, there is a marked gap between these ambitions and clinical reality. Studies indicate that older adults and individuals with limited health or digital literacy often face structural and educational barriers, which remain inadequately addressed despite the global policy focus [37,38]. Assessing digital health literacy is only the first step in identifying patients who may benefit from tailored support and in guiding the design of digital health tools that are accessible to patients with varying levels of digital competence.

Our exploratory analysis of the 3 digital health literacy subscales ("awareness," "skills," and "evaluate") indicates that patients may particularly struggle with critically appraising digital health information. Such limitations may increase the risk of misinterpreting online content, potentially affecting patient-physician communication and informed decision-making [39]. The assessment of separate domains in the 3-factor structure can reveal nuances that the overall score might hide [16]. Our findings suggest that targeted support to enhance evaluation skills may help patients engage more effectively and safely with digital health information.

Participants generally stated that it was "important" for them to have access to online health information and that the internet was "useful" in helping them make decisions regarding their health. This self-reported confidence likely reflects the well-established digital health infrastructure in the Danish health care system, which promotes patient engagement and access to personal health data [40]. Such a digital infrastructure likely contributes to patients' perceived ability to manage their health information online. This aligns with evidence suggesting that the use of internet-based health interventions can enhance digital health literacy [41].

Strengths and Limitations

The inclusion of 3 university hospitals covering approximately two-thirds of the national population is a major strength that enhances the generalizability of our findings. However, the exclusion of patients from the Capital Region, which is characterized by a younger and more highly educated population, may result in an underestimation of overall digital health literacy and limit the completeness of national representation.

Denmark's advanced digital infrastructure makes it a "critical case" for examining the integration of digital health solutions, offering valuable insights applicable to other countries navigating the digitalization of health care [40]. However, in settings with less integrated digital health systems or lower population-level digital competence, digital health literacy levels may be lower, and engagement with digital platforms may be more limited. Previous work has emphasized that differences in digital health literacy must be addressed to avoid inequities, highlighting the importance of simplifying communication and providing flexible solutions for disadvantaged populations [6]. The all-comer design, focusing specifically on elective open-heart surgery patients, further strengthens this study by targeting a group well positioned to benefit from tailored digital health interventions during recovery.

This study also had several limitations. First, the representativeness of the sample may be limited. Overall response rates were modest across all 3 sites, with particularly low participation at one site. Staff involved in questionnaire distribution reported that nonresponse was largely random due to logistical challenges, with only a minority attributable to patient exclusions or refusals. For context, data from the Western Denmark Heart Registry, which mandatorily records all cardiac surgery procedures in the catchment areas of the 3 participating regions, show that, for the same year, the mean age (65.9, SD 10.6 years) and proportion of male individuals

XSL•FO
RenderX

(76.9%) were nearly identical to those of our sample (mean age 65.8, SD 11 years; 232/311, 74.6%), supporting reasonable representativeness. However, selection bias cannot be entirely excluded. Excluding 22 patients with incomplete eHEALS questionnaires from the analysis may also have introduced selection bias.

Second, the questionnaire was administered on the day immediately preceding major surgery. Preoperative stress and anxiety could have influenced patients' self-reported digital health literacy, potentially inflating scores through social desirability or deflating them due to impaired concentration. However, as patients routinely receive electronic preoperative information, this timing mirrors real-world clinical practice.

Third, acute and subacute patients were not included; while their baseline digital health literacy remains unclear, their capacity to engage with digital health tools may be temporarily compromised due to the urgency of their conditions.

Finally, the use of the self-reported eHEALS, which does not fully capture the rapidly evolving digital health environment, may lead to an overestimation of patients' true digital health literacy levels.

Conclusions

This study demonstrates that patients scheduled for elective open-heart surgery generally report high digital health literacy, particularly younger individuals and those with higher levels of education. However, variability exists across sociodemographic groups, and challenges remain in patients' ability to critically evaluate online health information. The findings highlight the importance of health care professionals being aware of patients' varying levels of digital health literacy to provide appropriate support. Tailoring digital interventions based on patients' capabilities is essential to ensure equitable access and optimize the integration of digital health solutions into cardiac care.

Acknowledgments

The authors thank the patients and clinical staff at the 3 sites for their participation and assistance. They acknowledge statistician Julie Havbro Lund for support with statistical analyses. Editorial and coding support was provided by OpenAI's ChatGPT (GPT-4).

Funding

This work was supported by an independent grant from the Novo Nordisk Foundation (grant 0083480). The foundation had no role in the study design; data collection, analysis, and interpretation; or manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to patient confidentiality but are available from the corresponding author on reasonable request.

Authors' Contributions

RD, DSS, ISM, and TM designed and initiated the study. RD supervised the data collection at 2 of the sites, and BB and KHB supervised the collection of data at the third site. RD, DSS, and KHB entered data into REDCap (Research Electronic Data Capture). RD analyzed the data and drafted the manuscript under the supervision of ISM, TM, BB, KHB, and KAS. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Box plots and scatterplot presenting the distribution of eHealth Literacy Scale scores across gender, age, presence of close relationships, educational levels, cohabitation status, and number of health issues.
[DOCX File, 6175 KB - cardio_v10i1e83454_app1.docx]

References

1. Willems SH, Rao J, Bhambere S, Patel D, Biggins Y, Guite JW. Digital solutions to alleviate the burden on health systems during a public health care crisis: COVID-19 as an opportunity. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e25021. [doi: 10.2196/25021] [Medline: 34033575]
2. Seixas AA, Olaye IM, Wall SP, Dunn P. Optimizing healthcare through digital health and wellness solutions to meet the needs of patients with chronic disease during the COVID-19 era. Front Public Health 2021;9:667654. [doi: 10.3389/fpubh.2021.667654] [Medline: 34322469]
3. Frederix I, Solmi F, Piepoli MF, Dendale P. Cardiac telerehabilitation: a novel cost-efficient care delivery strategy that can induce long-term health benefits. Eur J Prev Cardiol 2017 Nov;24(16):1708-1717. [doi: 10.1177/2047487317732274] [Medline: 28925749]

XSL•FO
RenderX

4. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
5. Deshpande N, Arora VM, Vollbrecht H, Meltzer DO, Press V. eHealth literacy and patient portal use and attitudes: cross-sectional observational study. JMIR Hum Factors 2023 Jan 27;10:e40105. [doi: 10.2196/40105] [Medline: 36705947]
6. Gybel Jensen C, Gybel Jensen F, Loft MI. Patients' experiences with digitalization in the health care system: qualitative interview study. J Med Internet Res 2024 Apr 11;26:e47278. [doi: 10.2196/47278] [Medline: 38602748]
7. Wiesmüller F, Haag D, Sareban M, et al. Clinical, psychological, physiological, and technical parameters and their relationship with digital tool use during cardiac rehabilitation: comparison and correlation study. JMIR Mhealth Uhealth 2025 Apr 8;13:e57413. [doi: 10.2196/57413] [Medline: 40198893]
8. Zangger G, Bricca A, Liaghat B, et al. Benefits and harms of digital health interventions promoting physical activity in people with chronic conditions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 6;25:e46439. [doi: 10.2196/46439] [Medline: 37410534]
9. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PK. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]
10. Scheenstra B, Bongers BC, Broeders B, et al. Reasons and predictors of non-participation in a personalized digital prehabilitation care trial for patients undergoing elective cardiothoracic surgery. Interdiscip Cardiovasc Thorac Surg 2023 Aug 3;37(2):ivad123. [doi: 10.1093/icvts/ivad123] [Medline: 37486278]
11. Singh GK, Lee H, Kim LH, Daus GP. Promoting health literacy as an important initiative in reducing health disparities and advancing health equity. Int J Transl Med Res Public Health 2024;8:e008. [doi: 10.25259/IJTMRPH_58_2024]
12. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
13. Palm P, Brors G, Borregaard B, Rasmussen TB, Helmark C, Norekval T. Psychometric evaluation of the Danish version of the electronic health literacy scale (eHEALS) in patients treated with percutaneous coronary intervention (PCI): a cross-sectional validation study. Eur J Cardiovasc Nurs 2022 Jul 2;21(Supplement_1). [doi: 10.1093/eurjcn/zvac060.117]
14. Lee J, Lee EH, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021;23(11):e30644. [doi: 10.2196/30644] [Medline: 34779781]
15. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth Literacy Scale (eHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017 Feb 27;19(2):e53. [doi: 10.2196/jmir.5998] [Medline: 28242590]
17. Richtering SS, Hyun K, Neubeck L, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4. [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
18. Ramstad KJ, Brørs G, Pettersen TR, et al. eHealth technology use and eHealth literacy after percutaneous coronary intervention. Eur J Cardiovasc Nurs 2023 Jul 19;22(5):472-481. [doi: 10.1093/eurjcn/zvac087] [Medline: 36190843]
19. Khan D, Fjerbæk A, Andreasen JJ, Thorup CB, Dinesen B. Cardiac surgery patients' e-health literacy and their use of a digital portal. Health Educ J 2018;77(4):482-494. [doi: 10.1177/0017896918756435]
20. International Standard Classification of Education - ISCED. UNESCO. URL: https://www.uis.unesco.org/en/methods-and-tools/isced [accessed 2026-02-12]
21. Kronisk sygdom i befolkningen [Web page in Danish]. Kommunernes Landsforening. URL: https://www.kl.dk/analyser/analyser/social-sundhed-og-aeldre/kronisk-sygdom-i-befolkningen [accessed 2026-02-12]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Aug;18(3):91-93. [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
24. Danish Act on Research Ethics Review of Health Research Projects (Act no.1338 of September 1, 2020). Retsinformation. URL: https://www.retsinformation.dk/eli/lta/2020/1338 [accessed 2026-02-24]
25. Hua Z, Yuqing S, Qianwen L, Hong C. Factors influencing eHealth literacy worldwide: systematic review and meta-analysis. J Med Internet Res 2025 Mar 10;27:e50313. [doi: 10.2196/50313] [Medline: 40063939]
26. Melholt C, Joensson K, Spindler H, et al. Cardiac patients' experiences with a telerehabilitation web portal: implications for eHealth literacy. Patient Educ Couns 2018 May;101(5):854-861. [doi: 10.1016/j.pec.2017.12.017] [Medline: 29305064]
27. Ullrich G, Bäuerle A, Vogt H, et al. Digital health literacy and attitudes toward eHealth technologies among patients with cardiovascular disease and their implications for secondary prevention: survey study. JMIR Form Res 2025 Mar 19;9:e63057. [doi: 10.2196/63057] [Medline: 40106277]
28. Yang E, Chang SJ, Ryu H, Kim HJ, Jang SJ. Comparing factors associated with eHealth literacy between young and older adults. J Gerontol Nurs 2020 Aug 1;46(8):46-56. [doi: 10.3928/00989134-20200707-02] [Medline: 32936926]

XSL•FO
RenderX

29. Numbers and statistics. Agency for Digital Government. 2025. URL: https://en.digst.dk/numbers-and-statistics [accessed 2026-02-12]
30. Győrffy Z, Boros J, Döbrössy B, Girasek E. Older adults in the digital health era: insights on the digital health related knowledge, habits and attitudes of the 65 year and older population. BMC Geriatr 2023 Nov 27;23(1):779. [doi: 10.1186/s12877-023-04437-5] [Medline: 38012565]
31. Verweel L, Newman A, Michaelchuk W, Packham T, Goldstein R, Brooks D. The effect of digital interventions on related health literacy and skills for individuals living with chronic diseases: a systematic review and meta-analysis. Int J Med Inform 2023 Sep;177:105114. [doi: 10.1016/j.ijmedinf.2023.105114] [Medline: 37329765]
32. Lawrence K. Digital health equity. In: Linwood SL, editor. Digital Health [Internet]: Exon Publications; 2022. URL: https://pubmed.ncbi.nlm.nih.gov/35605078/ [accessed 2026-02-12]
33. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27. [doi: 10.2196/jmir.6709] [Medline: 28119275]
34. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
35. Van De Belt TH, Engelen L, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010 Jun 11;12(2):e18. [doi: 10.2196/jmir.1350] [Medline: 20542857]
36. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2026-02-12]
37. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
38. Arcury TA, Sandberg JC, Melius KP, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150. [doi: 10.1177/0733464818807468] [Medline: 30353776]
39. Vega ES, Zepeda MF, Gutierrez EL, Martinez MC, Gomez SJ, Caldera SD. Internet health information on patient's decision-making: implications, opportunities and challenges. Med Res Arch 2023;11(7.2). [doi: 10.18103/mra.v11i7.2.4066]
40. Eriksen J, Ebbesen M, Eriksen KT, et al. Equity in digital healthcare - the case of Denmark. Front Public Health 2023;11:1225222. [doi: 10.3389/fpubh.2023.1225222] [Medline: 37744503]
41. Wu KA, Kunte S, Rajkumar S, et al. Digital health for patients undergoing cardiac surgery: a systematic review. Healthcare (Basel) 2023 Aug 28;11(17):2411. [doi: 10.3390/healthcare11172411] [Medline: 37685445]

Abbreviations

eHEALS: eHealth Literacy Scale
REDCap: Research Electronic Data Capture
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Singh; submitted 03.Sep.2025; peer-reviewed by L Kayser, SCL Au; revised version received 14.Jan.2026; accepted 23.Jan.2026; published 27.Feb.2026.

Please cite as:
Daugaard R, Maribo T, Borregaard B, Skydt DS, Bech KH, Skovli KA, Modrau IS
Digital Health Literacy in Elective Open-Heart Surgery Patients: Cross-Sectional Study
JMIR Cardio 2026;10:e83454
URL: https://cardio.jmir.org/2026/1/e83454
doi:10.2196/83454

© Rikke Daugaard, Thomas Maribo, Britt Borregaard, Ditte Sommerlund Skydt, Kristina Hindhede Bech, Kjersti Alexandra Skovli, Ivy Susanne Modrau. Originally published in JMIR Cardio (https://cardio.jmir.org), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Novel Convolutional Neural Network–Based Algorithm for Heart Rate Measurement From Ballistocardiography Signals in Diverse Clinical Settings: Observational Study

Kumar Chokalingam; Muthukumarasamy Saravanan; Ashish Kaushal; Srishti Rao; Inam Ur Rahman; Ashwathi Nambiar; Mudit Dandwate; Ravi Mahajan; Kunal Sarkar; Gaurav Parchani

Department of Clinical Research, Turtle Shell Technologies Private Limited, Ground and Mezzanine Floor, Nomads Daily Huddle, City Centre, 40, Chinmaya Mission Hospital Road, Stage 2, Hoysala Nagar, Indiranagar, Bengaluru, India

Corresponding Author:
Kumar Chokalingam
Department of Clinical Research, Turtle Shell Technologies Private Limited, Ground and Mezzanine Floor, Nomads Daily Huddle, City Centre, 40, Chinmaya Mission Hospital Road, Stage 2, Hoysala Nagar, Indiranagar, Bengaluru, India

Abstract

Background: Continuous vital sign monitoring ensures early detection, prevents intensive care unit (ICU) admissions, and improves patient outcomes. Continuous heart rate (HR) monitoring methods often require direct skin contact, which can lead to patient discomfort. The rising popularity of ballistocardiography (BCG) offers a promising, noncontact solution for continuous vital sign monitoring with improved patient comfort.

Objective: This study aims to develop and validate a novel HR measurement algorithm leveraging convolutional neural networks (CNNs) and BCG signals for accurate, noncontact, and continuous HR monitoring. By integrating time-domain peak detection with short-time Fourier transform and CNN models, the proposed approach seeks to enhance HR measurement accuracy across diverse health care settings. The study follows the Food and Drug Administration (FDA)'s Good Machine Learning Practice guidelines and evaluates the algorithm's robustness, generalizability, and clinical applicability through extensive testing on a diverse dataset, ensuring improved patient comfort and early detection of clinical deterioration.

Methods: The proposed algorithm combines time-domain peak detection with short-time Fourier transform and CNNs to enhance HR measurement from BCG signals. The CNN model developed was trained on 129,976 data points from 373 participants (HR range: 36 - 230 bpm), including ICU patients, and was tuned on 75,970 data points from 192 participants (HR range: 46 - 169 bpm), with HR obtained from clinical-grade electrocardiography devices to improve generalizability. The algorithm was tested on 70,211 data points from 205 participants, including ICU patients, across 5 independent studies to demonstrate robust performance against diverse settings, demographics, and comorbidities. The methodology is in compliance with the FDA's *Good Machine Learning Practice for Medical Device Development: Guiding Principles*.

Results: The algorithm achieved a mean absolute error of under 3 bpm and a detection rate exceeding 80%, underscoring its robustness. The Bland-Altman analysis indicates high accuracy with a minimal bias of 0.25 and limits of agreement within 8.59 bpm. Additionally, the Pearson correlation coefficient of 0.97 from the Deming regression further demonstrates strong alignment with reference HR measurements, reinforcing its precision and reliability for clinical applications.

Conclusions: This CNN-based algorithm presents a robust solution for contactless HR monitoring, addressing the limitations of prior methods in noise management and adaptability. Its demonstrated accuracy, particularly in real-world, noisy clinical environments, highlights its potential for broad application in patient monitoring and improved comfort.

(JMIR Cardio 2026;10:e71224) doi:10.2196/71224

KEYWORDS

heart rate measurement; ballistocardiography; convolutional neural networks; contactless monitoring; health care technology; signal processing; remote patient monitoring

Introduction

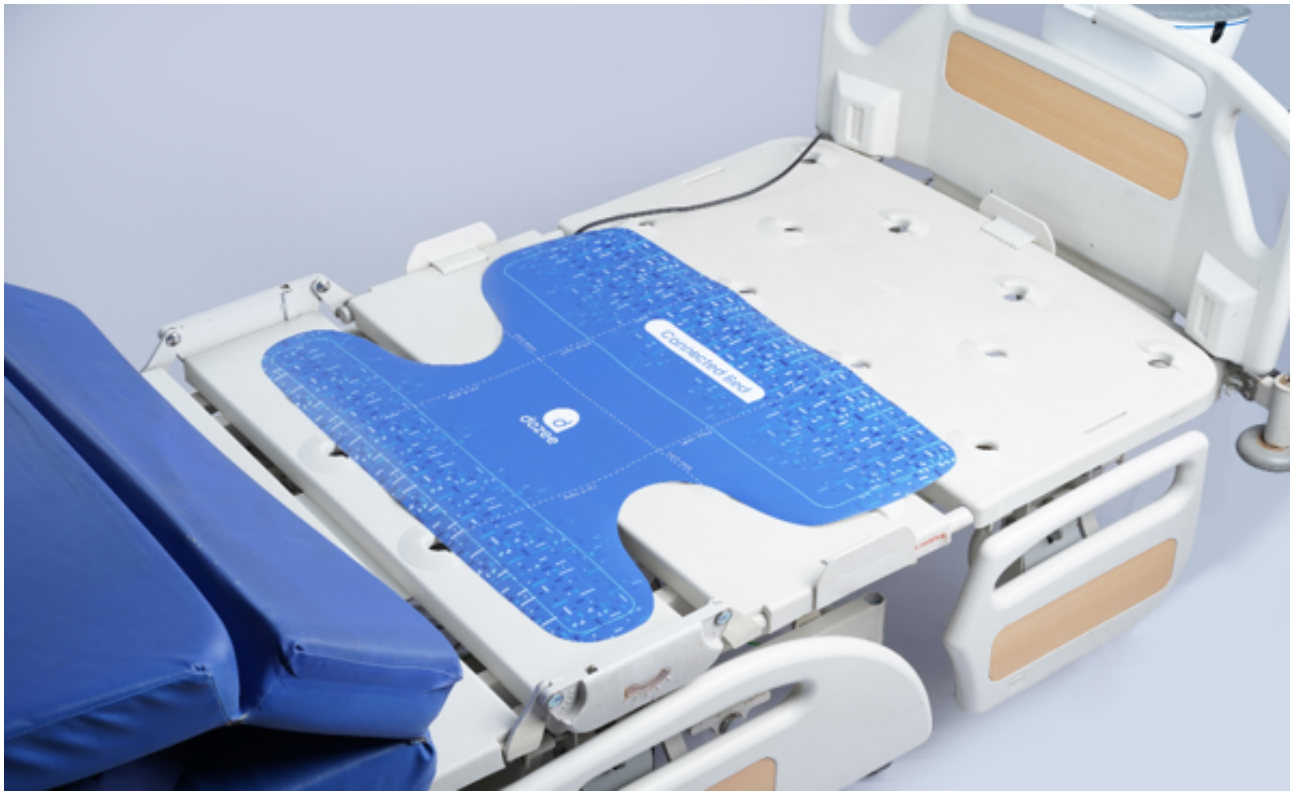
Continuous monitoring of heart rate (HR), respiratory rate, and blood pressure in ward patients enables early detection of deterioration, reducing unplanned intensive care unit (ICU) admissions and rapid response activations while improving outcomes [1-4]. Current continuous HR monitoring techniques, such as electrocardiography (ECG) [5] and photoplethysmography [6,7], require direct skin contact, which can cause discomfort and increase the risk of irritation during prolonged use [8,9]. This limitation is particularly relevant for patients requiring continuous monitoring, where skin-contact

XSL•FO
RenderX

sensors may further complicate patient care and comfort. As health care trends shift toward noninvasive solutions, there is a growing demand for continuous, contactless HR monitoring technologies that ensure patient comfort without compromising accuracy.

Ballistocardiography (BCG), which captures the body's micromovements induced by cardiac contractions, has emerged as a promising contactless technique for monitoring cardiovascular activity. By detecting mechanical vibrations generated by the heart's pumping action, BCG offers a noninvasive means of measuring HR [10-12]. This makes it particularly suitable for long-term monitoring in settings such as older adult care, postoperative recovery, and remote patient care, where minimizing discomfort and maximizing patient compliance are essential [13]. Dozee (Turtle Shell Technologies Pvt Ltd) has successfully integrated BCG technology with advanced algorithms to develop a remote patient monitoring system capable of continuously tracking HR, respiratory rate, and sleep patterns [3,14-19]. This system uses a sensor sheet positioned beneath the mattress (Figure 1) to capture vital signs without direct patient contact, providing an unobtrusive solution that is ideal for long-term use in various health care environments.

Figure 1. A sensor sheet positioned beneath the bed captures subtle microvibrations from the patient, enabling noninvasive measurement of vital signs.



Despite its potential, HR measurement from BCG signals remains challenging due to noise, motion artifacts, and physiological variability, which can degrade signal quality and compromise accuracy [20-23]. Reliable HR measurement requires isolating cardiac components from a composite signal that also includes respiratory information, body movement information, and environmental noise. To achieve this, BCG signal processing often uses band-pass filtering or advanced decomposition techniques to separate signal components. Following signal separation, various algorithmic approaches—such as time-domain, frequency-domain, and wavelet-transform methods—can be applied to extract meaningful HR metrics [24].

Previous BCG-based HR estimation methods relied heavily on traditional signal processing techniques, such as peak detection in the time domain. For instance, Dozee earlier approaches used time-domain peak detection, which demonstrated good performance in controlled conditions, with HR measurements typically within 3 bpm of reference values [15,18,19]. However, this accuracy diminished in real-world environments due to noise and body movement, limiting the effectiveness of these methods in clinical scenarios.

Recent advancements in machine learning (ML) and artificial intelligence have enabled the development of data-driven models that can recognize complex patterns in noisy datasets, providing significant improvements in HR measurement accuracy. These ML-based methods often require extensive datasets and substantial computational resources. However, many current ML-based models are trained on limited datasets, which can hinder their ability to generalize effectively across diverse patient populations and clinical conditions [25,26].

In response to these limitations, we introduce a novel convolutional neural network (CNN)-based algorithm designed to improve HR estimation from BCG signals. This algorithm combines time-domain peak detection with frequency-domain analyses, leveraging CNNs to extract higher-order features from the BCG signal. By using both time- and frequency-domain information, our hybrid approach effectively addresses challenges related to noise and motion artifacts, offering

XSL•FO
RenderX

enhanced accuracy in diverse and complex clinical environments.

Our algorithm was trained on a large, demographically diverse dataset encompassing various clinical and nonclinical environments to ensure robustness and generalizability. Following the practices laid down in the Food and Drug Administration (FDA)'s *Good Machine Learning Practice for Medical Device Development: Guiding Principles*, we tested the algorithm on a mutually exclusive dataset collected independently from the training and tuning data [27]. This rigorous validation underscores the algorithm's potential for reliable, contactless HR monitoring across a wide range of health care applications, from hospital wards to remote patient monitoring [27].

Methods

Study Design

This retrospective analysis used previously collected BCG signals and reference HR data from multiple research studies, including those in which HR evaluation was either a primary or secondary objective.

Ethical Considerations

For studies in which HR evaluation was not a primary or secondary focus, as well as for the overall retrospective analysis, ethical approval (ECG024/2024) was obtained from the Genebandhu Independent Ethics Committee, New Delhi, India, to ensure adherence to established ethical guidelines. Patient confidentiality was rigorously maintained through the anonymization of all data in compliance with data protection regulations. Furthermore, informed consent was obtained from all participants during the original data collection, upholding ethical standards and safeguarding participants' rights. No compensation was provided to participants.

Dataset Overview

The dataset includes BCG signals and reference HR measurements gathered from a range of clinical and nonclinical environments. In these studies, BCG and reference HR measurements were captured simultaneously, with HR data obtained from clinical-grade ECG devices. Reference HR values from clinical-grade ECG devices served as an accurate benchmark for evaluating the algorithm's performance. BCG data were gathered using the Dozee contactless monitoring system. Data collection covered a range of conditions, including rest, movement, and sleep. All reference ECG-derived HR values underwent quality screening to remove physiologically implausible readings or artifact-induced errors. Spurious values (eg, HR<20 or >250 bpm, or sudden changes >50% within 1 measurement interval) were flagged and excluded. By filtering out these outliers and any HR readings noted as affected by clinical interventions, we ensured that the reference standard was reliable for algorithm training and validation.

For the algorithm's development and validation, the dataset was segmented into 3 subsets—training, tuning, and testing—reflecting different stages of the algorithm's development. These subsets encompass a diverse participant pool, ensuring representation across various age groups, sex, and health conditions.

Notably, the training or tuning and testing sets were defined a priori by separate clinical researchers, and the data scientists developing the model were blinded to the allocation. The training or tuning and testing partitions were selected to ensure broad representation across HR ranges, comorbidities, demographics, and participant characteristics. This strict separation and blinding process followed recommended practices, eliminating any risk of data leakage between training and validation cohorts and ensuring an unbiased performance evaluation.

Training and Tuning Datasets

The training and tuning datasets were drawn from 10 studies conducted at 6 locations across India (Table 1), covering both clinical and home settings to support the comprehensive development of the algorithm across diverse environments and patient demographics. The training dataset included data from 373 participants, totaling 129,976 data points (Table 2). This dataset was used to train a CNN to recognize HR patterns within BCG signals, covering an HR range of 36 to 230 bpmTable 2 and Figure 2A) and a varied range of comorbidities (Figure 3A). This extensive range supports the algorithm's robustness across varied cardiac activities and demographics.

Table . Overview of study locations and participant populations for the training and tuning dataset.

Location	Number of studies	Study participants
1. Apollo Hospitals, Jubilee Hills, India	3	ICU[a] and ward patients
2. Sparsh Hospitals, Bengaluru, India	2	Healthy volunteers and ward patients
3. Jayadeva Institute of Cardiovascular Sciences and Research, Bengaluru, India	1	Ward patients
4. National Institute of Mental Health and Neurosciences, Bengaluru, India	2	Healthy volunteers
5. Indira Gandhi Institute of Medical Sciences, Patna, India	1	ICU patients
6. Dozee and Home Settings, Bengaluru, India	1	Healthy volunteers

[a]ICU: intensive care unit.

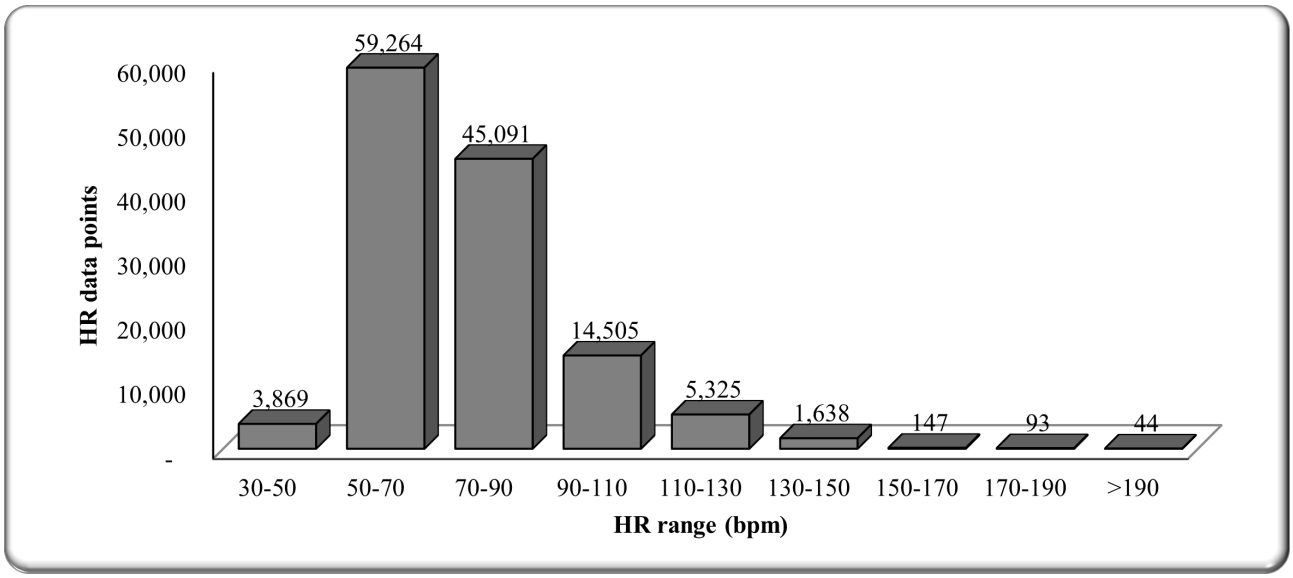
XSL•FO
RenderX

Table . Summary of demographic and heart rate (HR) data across datasets.

	Training	Tuning	Combined
Subjects	373	192	565
Sex (male:female)	242:128[a]	116:70[c]	358:198[e]
Age (y), mean (SD; range)	38[b] (16.83; 19-90)	37[c] (16.36; 19-88)	38[f] (16.66; 19-90)
Weight (kg), mean (SD; range)	69.16[b] (14.98; 35-125)	71.18[d] (16.22; 40-135)	69.83[g] (15.42; 35-135)
Height (m), mean (SD; range)	1.67[b] (0.10; 1.25-1.87)	1.68[e] (0.09; 1.49-1.85)	1.67[h] (0.10; 1.25-1.87)
BMI, mean (SD; range)	24.86[b] (4.93; 14.17-48)	25.27[e] (4.85; 16.22-43.70)	24.99[h] (4.90; 14.17-48)
HR data points	129,976	75,970	205,946
HR range	36-230	46-169	36-230

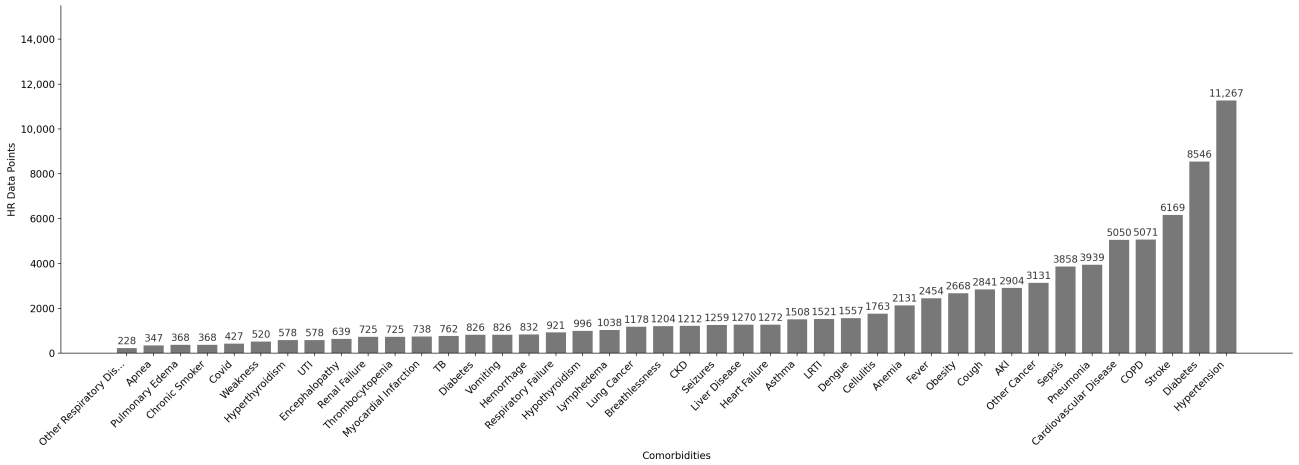
[a] = 3
[b] = 5
[c] = 6
[d] = 7
[e] = 9
[f] = 11
[g] = 12
[h] = 14

Figure 2. Heart rate (HR) data points across different ranges in the (A) training dataset and (B) tuning dataset showcasing a diverse representation across all ranges.



XSL•FO
RenderX

Figure 3. Heart rate (HR) data points across comorbidities in the (A) training dataset and (B) tuning dataset, showcasing a diverse representation of health conditions. AKI: acute kidney injury; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; LRTI: lower respiratory tract infection; TB: tuberculosis; UTI: urinary tract infection.



The tuning dataset included 75,970 data points from 192 participants Table 2 and was used to refine the algorithm's parameters for optimal performance. This subset facilitated adjustments to the CNN's architecture and hyperparameters, enhancing the algorithm's ability to generalize beyond the training data. The HR range within this tuning dataset, 46 to 169 bpm (Table 2 and Figure 2B), reflects typical cardiac activity in the general population, and the inclusion of a varied range of comorbidities (Figure 3B) further strengthens the algorithm's applicability across diverse user groups.

Testing Dataset

The testing dataset was sourced from 5 independent clinical studies conducted at various locations (Table 3), entirely separate from the training and tuning data to ensure unbiased evaluation and minimize the risk of overfitting. This dataset included data from 205 participants and 70,211 data points, covering an HR range of 40 to 169 bpm (Table 4 and Figure 4) and a range of comorbidities, each with at least 3000 data points (Figure 5). The diversity in participant demographics, including different age groups, sex, and health conditions, provided a robust foundation for evaluating the algorithm's performance across a wide range of patient populations and clinical environments.

Table . Overview of study locations and participant populations for the testing dataset.

Study	Study title	Location	Study participants
Study 1	Pilot Clinical Evaluation of Dozee VS in Hospital Patients	Eastside Research Centre, Seattle, United States	Healthy volunteers
Study 2	Evaluation of Continuous Non-Invasive Blood Pressure Monitoring through Dozee	BGS Global Gleneagles Hospital, Bengaluru, India	ICU[a] patients
Study 3	Dozee as a Screening Tool for Sleep Apnea Detection and Classification	St. John's Medical College Hospital, Bengaluru, India	Patients with apnea
Study 4	Dozee as a Screening Tool for Sleep Apnea Detection and Classification	Apollo Hospital, Bannerghatta Road, Bengaluru, India	Patients with apnea
Study 5	Dozee as a Screening Tool for Sleep Apnea Detection and Classification	Nithra Institute of Sleep Sciences, Chennai, India	Patients with apnea

[a]ICU: intensive care unit.

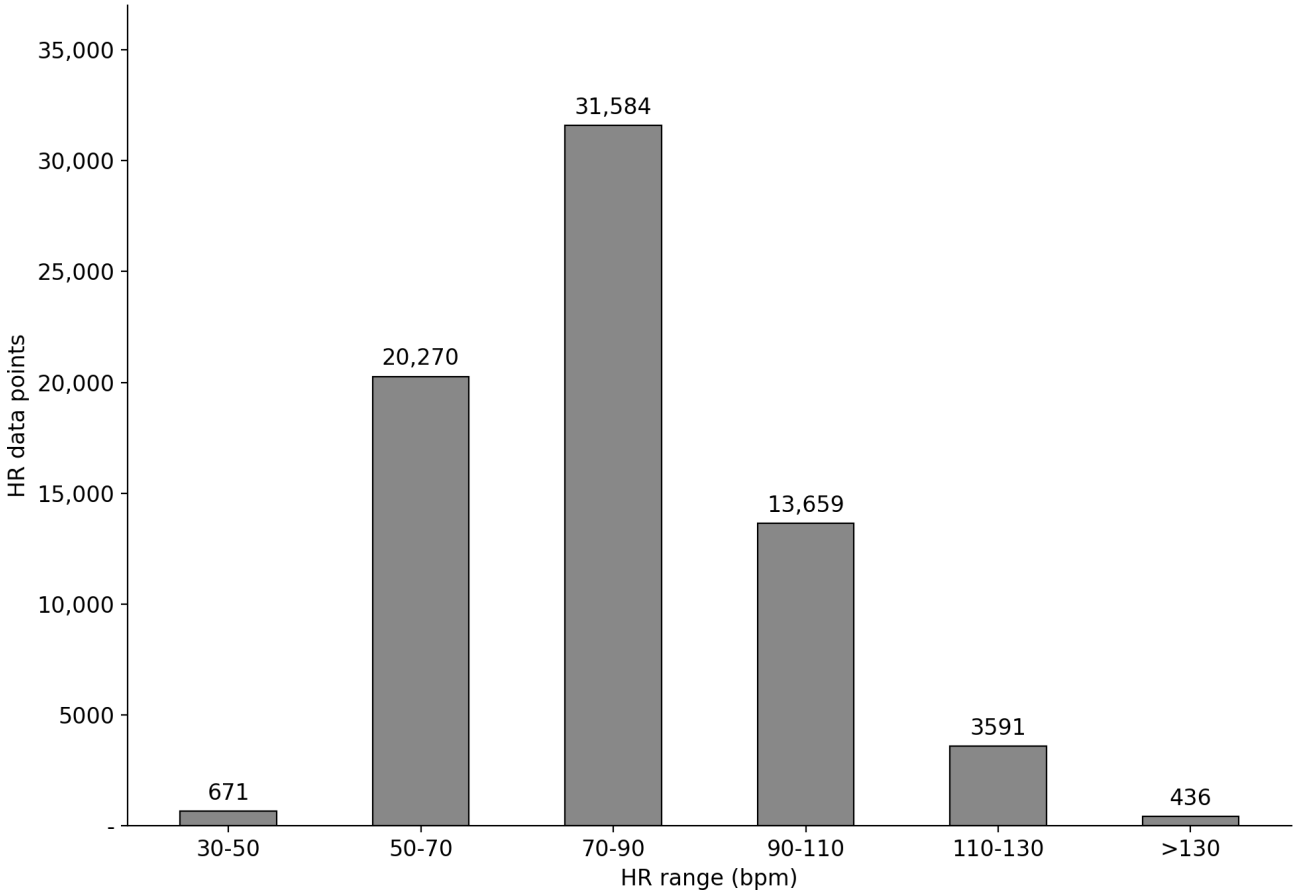
XSL•FO
RenderX

Table . Overview of demographics and heart rate (HR) for the testing dataset.

Characteristics	Study 1 (n=50)	Study 2 (n=71)	Study 3 (n=25)	Study 4 (n=53)	Study 5 (n=6)	Combined (n=205)
Sex (male:female)	12:38	47:23[a]	18:7	32:21	6:0	115:89[a]
Age (y), mean (SD; range)	33 (13.74; 18-58)	53 (17.19; 20-94)[a]	49 (15.19; 19-84)[a]	54 (14.47; 18-90)	45 (12.94; 27-59)	48 (17.47; 18-94)[a]
Weight (kg), mean (SD; range)	72.48 (15.85; 46-116)	69.39 (11.92; 50-115)[a]	81.19 (12.94; 55 - 103.5)[a]	86.28 (15.57; 59-132)	90.63 (27.29; 65-131)	76.58 (16.26; 46-132)[a]
Height (m), mean (SD; range)	1.68 (0.11; 1.49 - 1.93)	1.67 (0.06; 1.44 - 1.75)[a]	1.63 (0.08; 1.52 - 1.80)[a]	1.63 (0.10; 1.40 - 1.85)	1.70 (0.08; 1.59 - 1.79)	1.66 (0.09; 1.40 - 1.93)[a]
BMI, mean (SD; range)	25.52 (4.58, 18 - 35.81)	24.88 (4.90; 19.53 - 48.23)[a]	30.64 (5.56; 19.86 - 40.82)[a]	32.47 (4.72; 23.1 - 45.1)	30.73 (6.75; 23.9 - 41.8)	27.88 (5.90; 18 - 48.23)[a]
HR data points	6086	27,194	11,189	23,263	2479	70,211
HR range	40 - 108	44 - 169	42 - 128	48 - 119	46 - 112	40 - 169

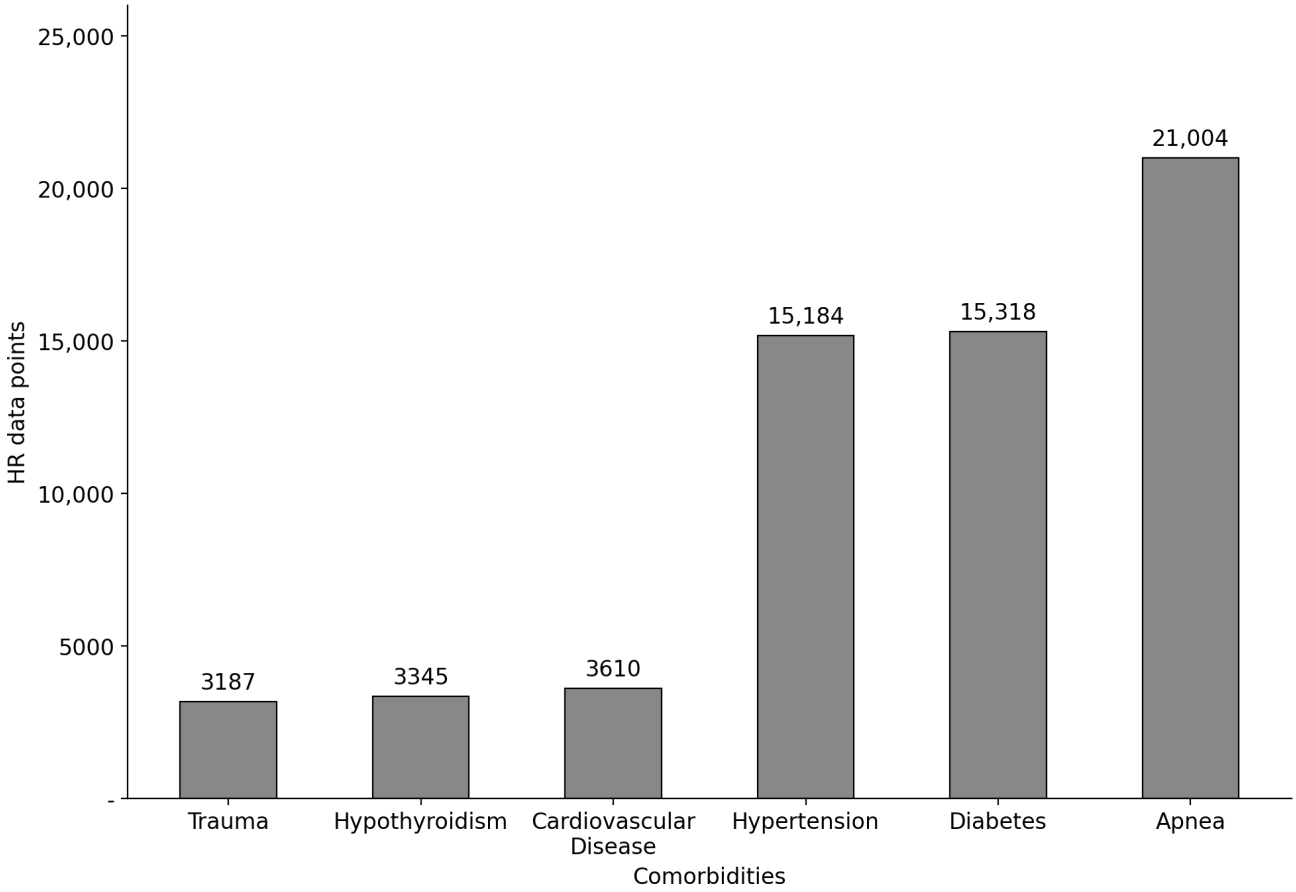
[a]Depicts the respective participants' missing data.

Figure 4. Heart rate (HR) data points across different ranges in the testing dataset, showcasing a diverse representation across all ranges.



XSL•FO
RenderX

Figure 5. Heart rate (HR) data distribution across comorbidities, each with at least 3000 data points.



Preprocessing of BCG Signals

Overview

Preprocessing steps were applied to BCG raw signals (Figure 6A and B, top panel) to reduce noise and artifacts introduced by movement and other physiological factors, ensuring a clean input for the HR measurement algorithm. Each 2-minute BCG recording was subsequently segmented into shorter epochs corresponding to periods of patient stillness. Segments contaminated by movement artifacts or excessive noise were automatically discarded, ensuring that only high-quality, artifact-free signal portions were used for HR analysis.

Artifact Removal

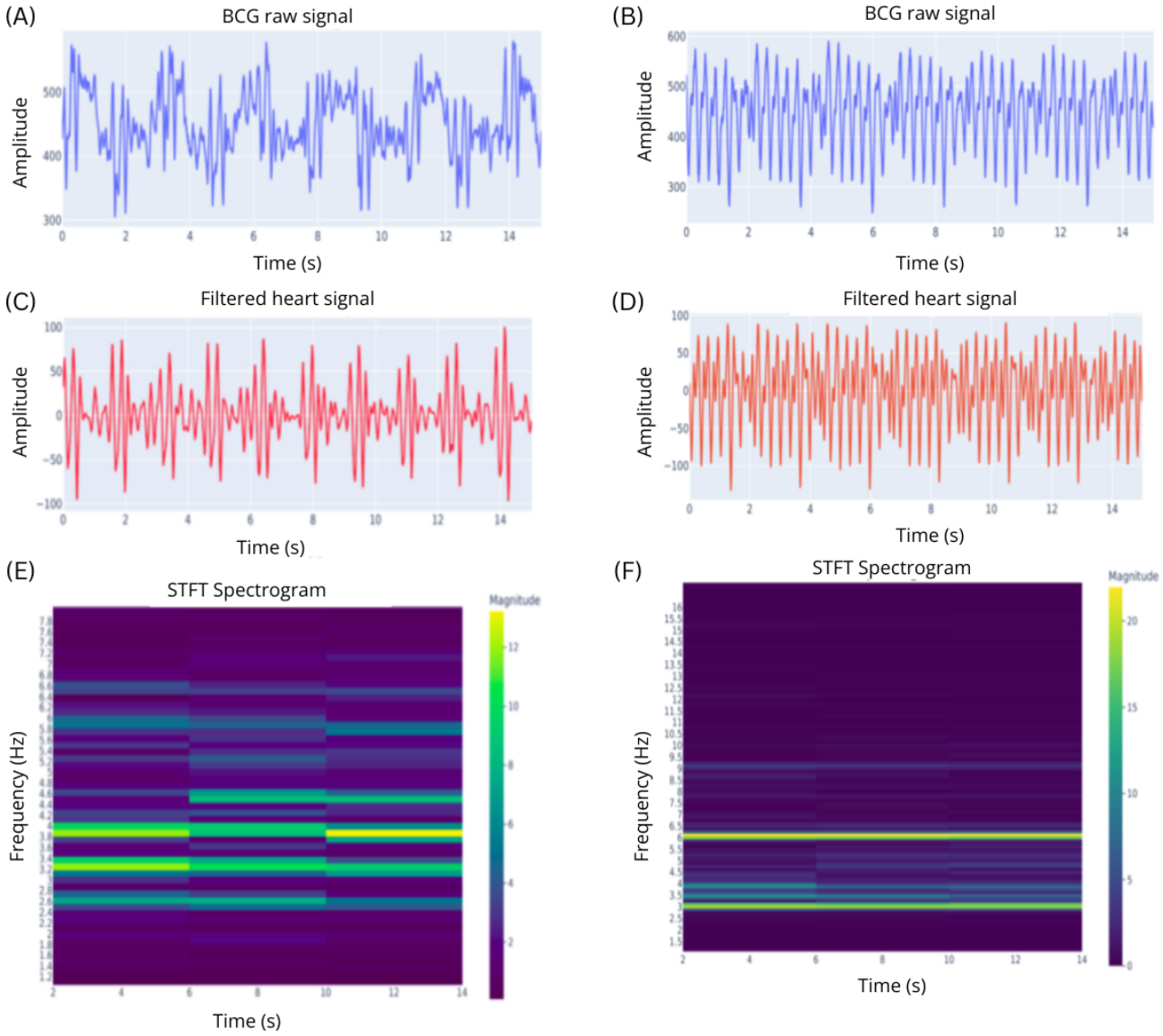
Motion artifacts were detected using adaptive filtering techniques, and corrupted signal segments were excluded from further analysis to improve the accuracy of HR measurement.

Noise Reduction

A band-pass filter of 0.5 to 25 Hz was applied to remove the low- and high-frequency noises and the respiration of BCG signals (Figure 6A and B, middle panel).

XSL•FO
RenderX

Figure 6. (A) & (B): Raw ballistocardiography (BCG) signals showing cardiac contractions and respiration. Filtered cardiac signals: (C) 9.5 beats in 15 seconds (38 bpm; reference 39 bpm) and (D) 46 beats in 15 seconds (182 bpm; reference 182 bpm). (E) & (F): Spectrograms highlighting the dominant cardiac frequencies for each segment. STFT: short-time Fourier transform.



HR Measurement Algorithm

The HR measurement algorithm was designed to extract fundamental and harmonic frequency components from the BCG signal, combining signal processing techniques with a CNN for robust HR detection under diverse conditions. Figure 6 illustrates the algorithm's process on a sample segment: the raw BCG waveform and its filtered version, alongside the short-term Fourier transform (STFT) spectrogram. In these examples, the algorithm correctly identifies an HR of approximately 39 bpm and approximately 182 bpm from BCG signals, closely matching the reference ECG. The HR measurement algorithm includes the following processes:

- Heart signal fundamental and harmonic frequencies extraction: a STFT was applied to the filtered signal segments, generating a spectrogram in which the periodic nature of heartbeats manifested as fundamental and harmonic frequency components. These frequencies, indicative of the regularity of heartbeats, were used to distinguish and infer the HR from other noises. For example, if the harmonic spacing corresponded to a fundamental frequency of 3.03 Hz, this translates to an HR of approximately 182 bpm (Figure 6A and B bottom panel).
- A CNN algorithm was trained with both BCG signals and spectrograms as input. The CNN learns to recognize HR patterns from the BCG signals and frequency components within the spectrogram. During inference, the CNN detects fundamental frequencies and counts heartbeats from BCG signals, providing accurate HR measurements even under challenging conditions such as noise or motion artifacts. The CNN architecture was specifically designed to handle complex patterns in BCG signals and their spectrograms, achieving high accuracy by simultaneously analyzing both frequency and time domains.
- Adaptive filtering: the approach uses an iterative, dynamically adjusted thresholding mechanism that continuously evaluates real-time signal quality, enabling the system to automatically refine its sensitivity and thresholds to distinguish genuine cardiac signals from transient artifacts and environmental noise.

XSL•FO
RenderX

Algorithm Evaluation

The performance of the HR measurement algorithm was assessed on the testing dataset by comparing measured HR values to reference values using the following metrics to evaluate accuracy and agreement:

- Mean absolute error (MAE): the average absolute difference between the measured and reference HR values was calculated, providing an indicator of the algorithm's overall accuracy.
- Detection rate (DR): the percentage of time the algorithm successfully detected and measured HR from BCG signals was evaluated to assess robustness in real-world clinical environments where signal quality may vary.
- Bland-Altman: this method was used to analyze the agreement between measured and reference HR values, identifying any systematic bias by examining the distribution of differences.
- Deming regression with the Pearson correlation coefficient: this regression method, which accounts for measurement errors in both variables, was used to evaluate the alignment between measured and reference HR values. The Pearson correlation coefficient was calculated to quantify the strength and direction of the linear relationship between the algorithm's HR measurements and clinical reference values.

These metrics were chosen to facilitate comparison with prior studies. MAE, Bland-Altman limits of agreement, correlation coefficients, and detection rates are standard benchmarks for validating unobtrusive HR estimation algorithms [28-34]. Using this common set of metrics allows us to contextualize our results against existing BCG-based and other noncontact HR monitoring approaches.

These evaluation metrics provide a comprehensive analysis of the algorithm's accuracy, reliability, and applicability in real-world clinical settings, reinforcing its potential as a contactless solution for HR monitoring.

Results

MAE and DR Analysis

The HR measurement algorithm demonstrated high accuracy and reliability across datasets from studies conducted in both the United States and India. In total, the algorithm processed 70,211 reference points and produced 58,908 measured points, resulting in an MAE of 2.50 bpm and a DR of 83.90%.

In the US-based study 1, the algorithm achieved an MAE of 2.27 bpm and a DR of 92.34%. The algorithm's performance remained robust in the India-based studies (studies 2 - 5), with a combined MAE of 2.52 bpm and a DR of 83.10%. Specifically, study 2 reported an MAE of 2.73 bpm and a DR of 81.40%, while study 3 achieved an MAE of 2.49 bpm and a DR of 81.12%. Study 4 achieved an MAE of 2.39 bpm and a DR of 84.52%. Notably, study 5 demonstrated the best performance, with an MAE of 1.78 bpm and a DR of 97.26%.

These findings highlight the algorithm's consistency across diverse geographic locations and clinical environments, maintaining an MAE below 3 bpm and a DR exceeding 80% across all datasets. Table 5 provides a detailed summary of the performance metrics from the five studies, offering a comprehensive overview of the algorithm's accuracy and detection capabilities.

Table . Summary of heart rate (HR) measurement algorithm performance across 5 studies.

Variables	Reference data points	Measured data points	MAE[a]	DR[b]
Total	70,211	58,908	2.50	83.90
Study wise				
Study 1	6086	5620	2.27	92.34
Study 2	27,194	22,137	2.73	81.40
Study 3	11,189	9077	2.49	81.12
Study 4	23,263	19,663	2.39	84.52
Study 5	2479	2411	1.78	97.26
Study location				
The United States	6086	5620	2.27	92.34
India	64,125	53,288	2.52	83.10

[a]MAE: mean absolute error.

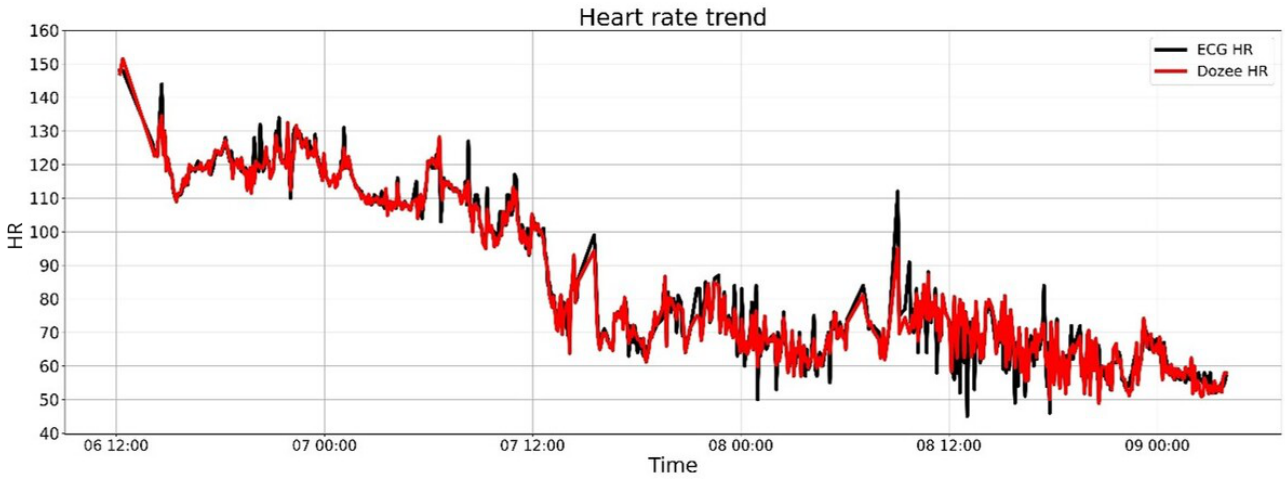
[b]DR: detection rate.

Figure 7 shows the HR trend for an ICU patient, comparing HR values measured by the CNN-based algorithm with clinical-grade ECG reference HR measurements over several hours. The figure illustrates that the HR measurement algorithm closely aligns with the ECG reference values, even as the patient's HR fluctuates within a broad range of 50 to 150 bpm. The visual alignment between the Dozee HR measurements and the ECG HR reference across varying HR levels highlights the algorithm's precision and responsiveness in a real-world clinical setting. This result demonstrates the algorithm's capability to maintain accurate HR tracking in the high-noise ICU

XSL•FO
RenderX

environment, underscoring its potential as a viable, contactless solution for real-time HR monitoring in health care settings.

Figure 7. Heart rate (HR) trend for an intensive care unit (ICU) patient, comparing HR values measured by the convolutional neural networks (CNNs)-based algorithm with clinical-grade electrocardiography (ECG) reference HR measurements over several hours.



A demographic analysis of the algorithm's performance showed stable accuracy across a variety of subgroups, including different HR ranges, age groups, sex, BMI categories, racial backgrounds, and comorbidities. Table 6 provides a summary of the algorithm's performance across these demographic variables.

XSL•FO
RenderX

Table . Summary of heart rate (HR) measurement algorithm performance across different demographic subgroups.

Demographic variables	Reference data points	Measured data points	MAE[a]	DR[b] (%)
HR (bpm)				
30 - 50	671	632	2.36	94.19
50 - 70	20,270	17,371	2.12	85.7
70 - 90	31,584	26,105	2.67	82.65
90 - 110	13,659	11,381	2.68	83.32
110 - 130	3591	3054	2.5	85.05
>130	436	365	2.36	83.72
Age (y)				
18 - 25	6290	5784	2.35	91.96
25 - 35	6706	5716	2.8	85.24
35 - 45	11,188	9678	2.46	86.5
45 - 55	17,572	14,920	2.21	84.91
≥55	27,967	22,508	2.64	80.48
Sex				
Female	26,466	21,943	2.77	82.91
Male	43,617	36,914	2.33	84.63
BMI (kg/m²)				
<18.5	121	121	1.52	100
18.5 - 25	23,791	19,383	2.59	81.47
25 - 30	19,569	17,134	2.5	87.56
30 - 35	17,882	15,205	2.36	85.03
≥35	8360	6763	2.48	80.9
Race				
Asian American	1122	1076	1.94	95.9
Black or African American	374	337	2.12	90.11
Indian	64,125	53,288	2.52	83.1
Mixed	244	210	2.71	86.07
Unknown or not reported	238	238	1.57	100
White or Caucasian	4108	3759	2.4	91.5
Study participants				
Healthy volunteers	6086	5620	2.27	92.34
ICU[c] patients	27,194	22,137	2.73	81.40
Patients with apnea	36,931	31,151	2.37	84.35
Comorbidities				
Hypertension	15,184	13,502	2.06	88.92
Diabetes	15,318	13,085	2.46	85.42
Cardiovascular disease	3610	2884	2.6	79.89
Apnea	21,004	17,305	2.68	82.39
Hypothyroidism	3345	2810	2.39	84.01
Trauma	3187	2864	1.73	89.87

[a]MAE: mean absolute error.

XSL•FO
RenderX

[b]DR: detection rate.

[c]ICU: intensive care unit.

The demographic subgroups are as follows:

- HR Range: the algorithm maintained an MAE of 3 bpm or lower across all HR ranges. Notably, in the lower HR range (30 - 50 bpm), the algorithm achieved a high DR of 94.19%, while for HRs above 130 bpm, the DR remained robust at 83.72%.
- Age: across all age categories, the algorithm showed strong performance with an MAE consistently below 3 bpm and a DR exceeding 83%, demonstrating its adaptability to different age groups.
- Sex: the algorithm performed equally well for both female and male participants, with MAE values of 2.77 bpm and 2.33 bpm, respectively, and similar DRs (82.91% for females and 84.63% for males).
- BMI: participants with a BMI of less than 18.5 kg m^{-2} achieved an MAE of 1.52 bpm and a DR of 100%, indicating particularly high accuracy in this group. For other BMI categories, the MAE remained below 3 bpm with detection rates above 80%.
- Race: performance was consistently high across racial backgrounds, with an MAE of 2.52 bpm and a DR of 83.1% for Indian participants, while Caucasian and African American groups also demonstrated high accuracy, with DRs exceeding 90%.
- Study participants: the algorithm demonstrated high accuracy across study groups, achieving an MAE of 2.27 bpm and a DR of 92.34% for healthy volunteers, while maintaining strong performance in ICU patients (MAE: 2.73 bpm, DR: 81.40%) and patients with apnea (MAE: 2.37 bpm, DR: 84.35%), showcasing its adaptability to diverse physiological conditions.
- Comorbidity: performance remained consistent across comorbidities, with an MAE below 3 bpm for most conditions. Analysis focused on comorbidities with at least 3000 data points, ensuring robust evaluation. Hypertension (MAE: 2.06 bpm; DR: 88.92%) and diabetes (MAE: 2.46 bpm; DR: 85.42%) showed strong performance, while trauma cases achieved the highest precision (MAE: 1.73 bpm; DR: 89.87%). Cardiovascular disease reported an MAE of 2.6 bpm and a DR of 79.89%, demonstrating reliability even in critical conditions.

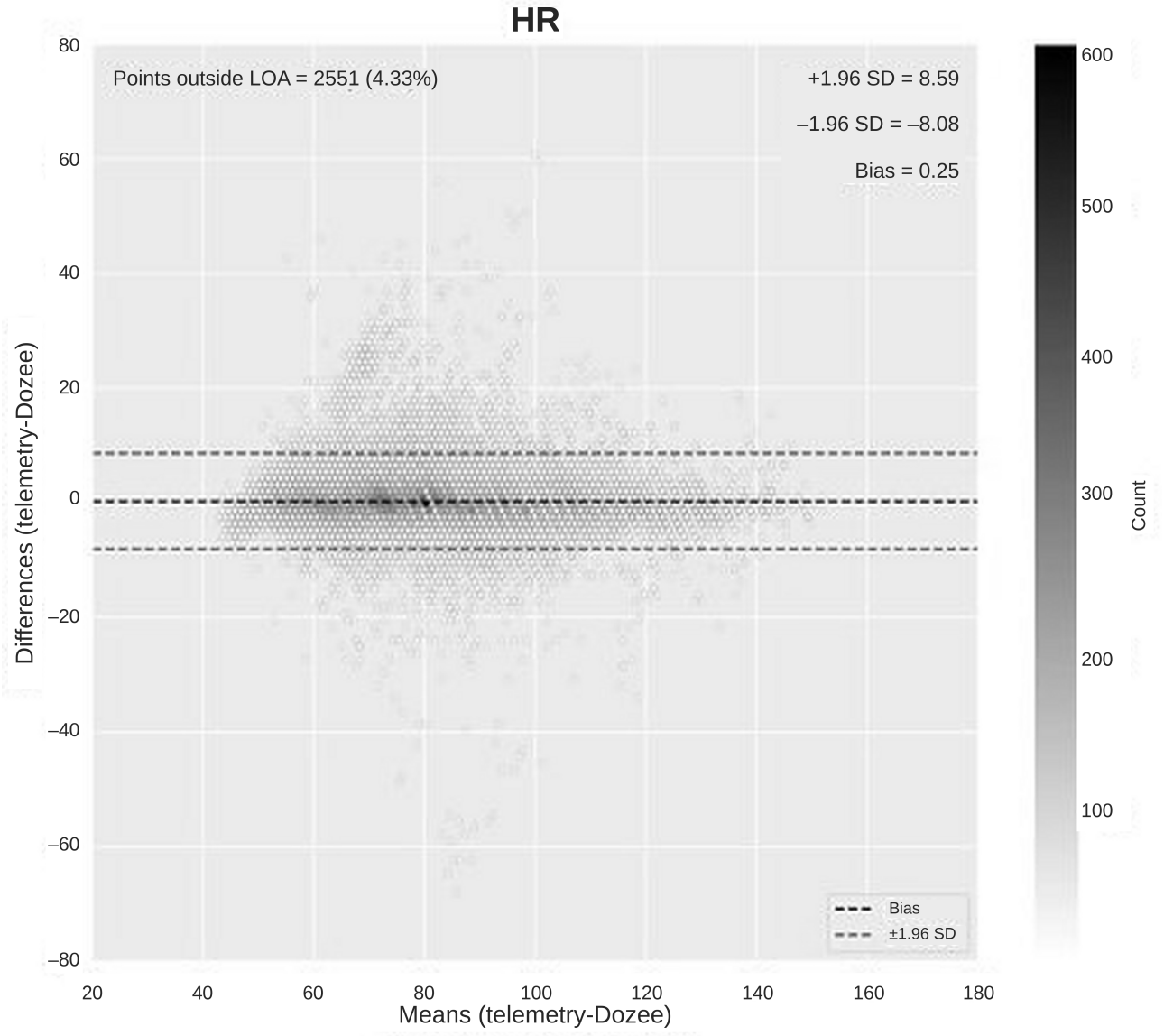
These findings confirm the algorithm's adaptability to diverse population profiles, underscoring its potential as a reliable HR monitoring tool across broad demographic groups in clinical settings.

Bland-Altman Analysis

The Bland-Altman analysis, shown in Figure 8, demonstrates the algorithm's high accuracy and reliability. The analysis revealed a minimal bias, with a point estimate of 0.25, indicating close alignment with reference HR values. The limits of agreement (LOA), ranging from 8.59 to −8.08, indicate that most differences between the measured and reference values are within 9 bpm, affirming the algorithm's precision. Notably, 2551 (4.33%) points fall outside the LOA, which represents a small proportion of the total dataset and underscores the algorithm's strong overall performance. These narrow LOA boundaries, combined with the low bias value, collectively highlight the algorithm's strong agreement with reference data, confirming its reliability for HR estimation in clinical environments where accuracy is essential.

XSL•FO
RenderX

Figure 8. Bland-Altman analysis of the entire testing dataset. HR: heart rate; LOA: limits of agreement.

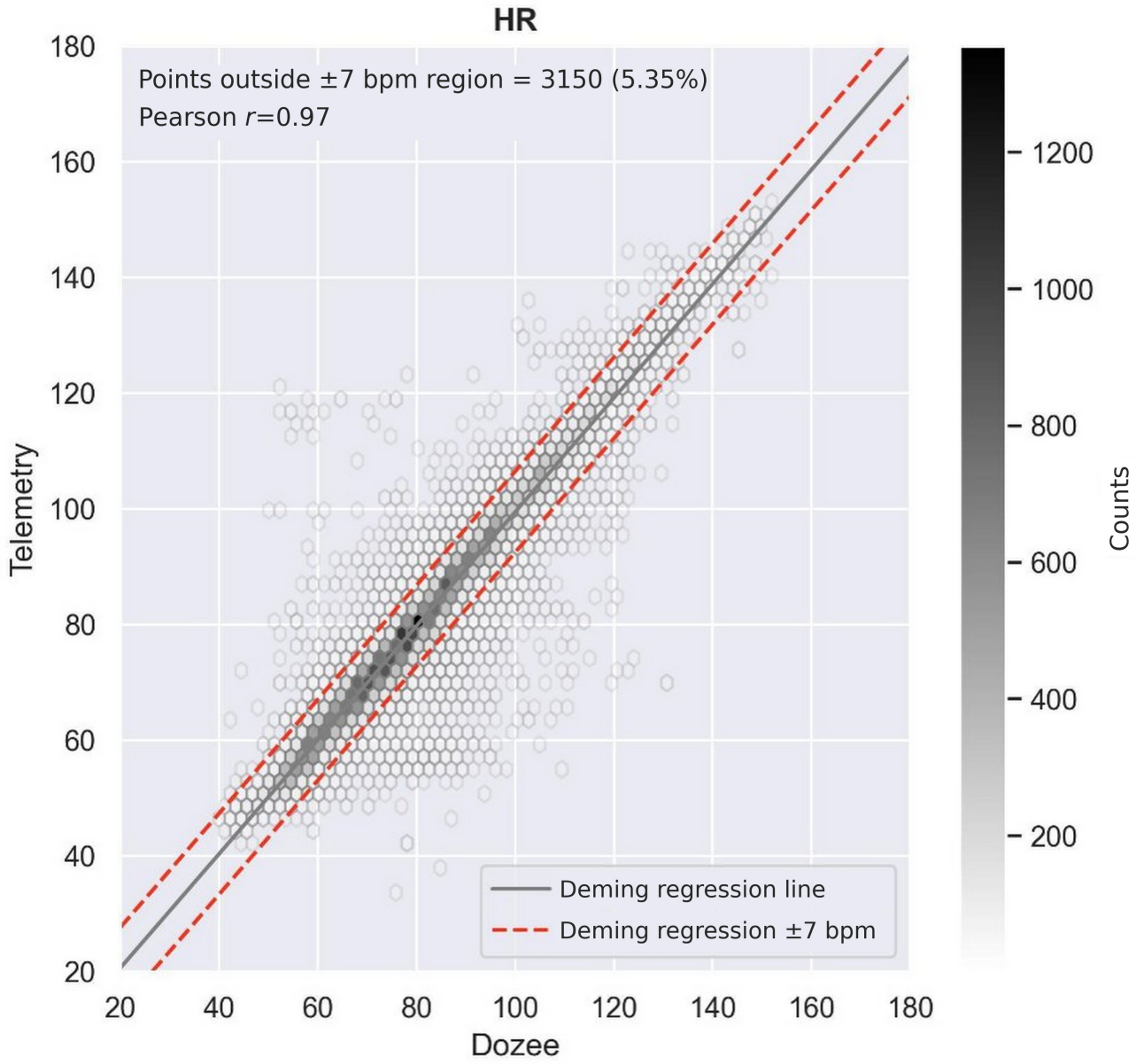


Deming Regression and Pearson Correlation Coefficient Analysis

The Deming regression analysis, complemented by the Pearson correlation coefficient, provided further evidence of the algorithm's alignment with reference HR values. Figure 9 illustrates the linear relationship between the algorithm's HR measurements and the clinical reference values. With a Pearson correlation coefficient of 0.97, the algorithm demonstrated a strong correlation with the reference HR data, reinforcing its accuracy and reliability. Notably, 3150 (5.35%) points fall outside the ±7 bpm region, further highlighting the precision of the algorithm's measurements. The high correlation, together with minimal observed bias, underscores the algorithm's potential for consistent and accurate HR monitoring in diverse clinical settings. These findings validate the algorithm's performance, confirming its suitability for contactless patient monitoring applications that require reliable HR data.

XSL•FO
RenderX

Figure 9. Deming regression with the Pearson coefficient for the entire testing dataset. HR: heart rate.



Discussion

Advancements in Contactless HR Monitoring

This study introduces a highly accurate, contactless HR measurement algorithm based on BCG signals, specifically designed for continuous, noninvasive monitoring in clinical settings. By combining adaptive signal processing techniques with ML-based peak detection, the algorithm effectively addresses challenges in BCG monitoring, such as noise, motion artifacts, and physiological variability. Unlike traditional methods that rely on fixed filtering and thresholding, this algorithm dynamically suppresses noise while preserving HR-related features, ensuring reliable performance under nonstationary conditions caused by patient movement and physiological fluctuations. The algorithm achieved an MAE within 3 bpm and a DR above 80% across diverse settings, demonstrating reliable HR monitoring despite motion artifacts and physiological variability.

A significant strength of this algorithm is its adaptability across demographic subgroups, including age, sex, BMI, and race. This versatility enhances its applicability in dynamic, real-world environments. The CNN architecture represents a substantial advancement over conventional threshold-based methods, which can obscure true cardiac peaks in noisy settings. By maintaining independent datasets for training, tuning, and testing, we minimized the risk of overfitting, ensuring the algorithm generalizes well across different environments. This careful design enables the algorithm to consistently deliver reliable HR measurements, making it suitable for hospital wards and potential applications in home and remote monitoring.

To our knowledge, this is the first BCG-based HR monitoring study to combine time-domain and frequency-domain processing with a deep learning model and to validate it across such a broad range of conditions. Our CNN-based algorithm demonstrates notable improvements in accuracy, robustness, and adaptability compared to existing BCG-based HR measurement solutions. With an MAE of 3 bpm, our algorithm surpasses the accuracy

XSL•FO
RenderX

of some commercial contactless solutions, such as EarlySense, which reports an MAE of 5 bpm, setting a new benchmark in contactless HR monitoring. Few studies report comparable MAE values. For example, Vesterinen et al [28] assessed the Emfit QS, a commercial contact-free BCG device, achieving an MAE of 2.4 bpm but with a higher rate of missing data compared to ECG. Similarly, Clemente et al [29] used a bed-mounted sensor system with an MAE of 2.41 bpm, while Proll et al [30] achieved an average MAE of 2.07 bpm across a smaller sample size of 14 patients. Jung et al [31] tested a multichannel load cell bed, achieving an MAE of 1.76 bpm in the supine posture, with a maximum error of 3.03 bpm in lateral postures. Although these solutions show promising accuracy, their limited validation in diverse clinical environments restricts generalizability.

The Bland-Altman analysis underscores the algorithm's high accuracy and reliability, validating its potential for clinical application. The minimal bias observed indicates a close alignment between the algorithm's HR measurements and the reference values, highlighting its precision and robustness under diverse conditions. The LOA show that most differences between measured and reference HR values fall within a clinically acceptable range, with only a small fraction of data points outside these limits. This reinforces the algorithm's reliability and strong agreement with reference data, even in the face of physiological and environmental variability. Notably, the bias and LOA observed in this analysis are comparable to those reported for HR measurements using BCG under real-world conditions. For instance, studies using BCG-based technologies, such as EarlySense in postoperative settings and Emfit in sleep lab environments [32,33], have demonstrated similar bias (−1.4 and 1.79) and LOA (10.4 to −13.2 and 19.11 to −15.33), further validating the robustness and reliability of the algorithm in practical, uncontrolled scenarios. This alignment with real-world performance benchmarks emphasizes the algorithm's suitability for clinical applications where accuracy and consistency are paramount.

The Deming regression analysis, supported by the Pearson correlation, highlights the strong alignment between the algorithm's HR measurements and the reference standard. The analysis demonstrates a robust linear relationship, reflecting the algorithm's precision and reliability across diverse conditions. This strong correlation is particularly important in clinical settings, where accurate trend detection and reliable HR measurements are essential for timely interventions and effective patient management. By ensuring minimal variability and maintaining close agreement with reference values, the algorithm enables health care providers to monitor physiological changes confidently. The findings align with those of other noninvasive HR monitoring systems [34], further validating the algorithm's performance. This establishes the algorithm as a viable tool for both hospital and remote monitoring, offering a reliable solution for continuous, contactless HR monitoring to improve patient care and outcomes.

Many existing approaches to HR estimation exhibit limitations in sample diversity and applicable HR ranges, which can impact their generalizability in clinical practice. For instance, Sumali et al [35] used a CNN algorithm for peak detection but limited their analysis to an HR range of 60 to 100 bpm, potentially overlooking extreme HRs encountered in clinical practice. Feng et al [36] used STFT and empirical mode decomposition for peak extraction on a small sample of 10 healthy participants and 2 patients with coronary heart disease, which limited applicability due to narrow demographics. Pino et al [37] tested only 34 normal participants and 24 with atrial fibrillation in controlled settings, which restricts the generalizability of their findings. Moukadem et al [38] based their analysis on data from just 3 participants, further limiting the applicability of their results. Yang et al [22] used peak detection with discrete wavelet transform across a small dataset of 24 participants, with recordings lasting about 3 minutes; this restricted duration and participant diversity limit applicability. Schranz et al [39] introduced a supervised deep learning method for J-peak extraction from BCG measurements. By modeling discrete heartbeat events with a symmetric, continuous kernel function (surrogate signal), this approach enhances target heartbeat detection. Their evaluation is based on a BCG dataset from 11 participants, collected over 17 nights of sleep, totaling 134 hours of ECG reference data that investigated various deep learning architectures, including convolutional, recurrent, and hybrid models, on a held-out testing set of 14 patients [37,39]. Choe et al [40] proposed an advanced algorithm using a moving dispersion calculation method to accurately identify heartbeat locations. This is enhanced by an adaptive peak detection technique that dynamically adjusts the detection window based on predicted heartbeat locations. The method effectively operates across a HR range of 40 to 110 bpm. In contrast to the above-mentioned studies, our algorithm stands out due to its validation across a broad HR range and diverse patient demographics, enhancing robustness and generalizability for real-world applications [40].

Currently, patients in hospital wards are often monitored solely through spot checks [41], highlighting the need for continuous HR monitoring. In this context, an error margin of 3 bpm is deemed acceptable for several reasons. This margin emphasizes trend detection over precise measurements, enabling health care providers to identify significant changes in HR that may indicate patient deterioration or response to treatment. Recognizing the natural fluctuations in HR due to physiological variability, this margin accommodates normal variations while still delivering clinically relevant data. Furthermore, it has been reported in the literature that HR is deemed clinically acceptable if it aligns within ±5 bpm or within ±10% of the reference standard, supporting the adequacy of these thresholds. Moreover, previous studies validate the adequacy of a ±3 bpm for effective monitoring, reinforcing its appropriateness for enhancing patient management and outcomes [3,42].

The key strength of this study lies in its innovative dual-domain algorithm, which integrates time-domain HR peak detection with STFT analysis. While the STFT effectively captures the fundamental heart-rate frequencies, the CNN adds significant value by identifying complex temporal and spectral patterns that may not be apparent in traditional frequency-domain representations and by improving robustness to noise and artifacts through learned pattern recognition. This is particularly advantageous in real-world clinical settings, where physiological variability and movement-related disturbances are common.

XSL•FO
RenderX

This dual-domain CNN approach addresses the noise or artifact challenges noted in earlier methods and elevates the performance to a level suitable for demanding health care environments.

By combining these advanced techniques, our algorithm sets a new benchmark in contactless HR monitoring. Rigorously trained on a large and diverse dataset that includes varying levels of motion artifacts and physiological diversity, particularly from ICU settings with inherent noise and HR fluctuation, this algorithm ensures reliable HR detection across a wide range of clinical scenarios, positioning it as a transformative solution for HR monitoring in challenging health care environments. Importantly, we validated the algorithm in both controlled and real-world environments, demonstrating practical robustness. Performance remains consistent across age, sex, BMI, ethnicity, and comorbidities, underscoring its strong generalizability.

While this study validates the algorithm effectively, several limitations must be considered. The BCG data used for training and validation were exclusively collected from piezoelectric sensors, which may limit the algorithm's applicability to other sensor types and reduce its generalizability. Testing across various BCG sensor technologies is recommended to enhance adaptability. Furthermore, while the algorithm performs well across a broad HR range, its accuracy with highly irregular heart rhythms, such as severe arrhythmias, remains untested and may require tailored adaptations. The CNN-based architecture also introduces computational demands that could hinder real-time processing in resource-limited environments; optimizing the algorithm for lower computational needs could improve scalability for telemedicine and home care applications. Additionally, we acknowledge a limitation in methodological transparency: the full CNN architectural details and certain parameter values are not disclosed due to the proprietary nature of the algorithm. This constraint, while necessary from an intellectual property standpoint, means that external researchers cannot yet fully reproduce the model. We have mitigated this by rigorously validating the algorithm on independent datasets. Finally, the algorithm's performance is influenced by the characteristics of the training data, potentially affecting generalization to demographics or conditions that are underrepresented in the dataset.

Future studies could further enhance the algorithm by integrating additional vital signs, such as respiratory rate, to create a more comprehensive monitoring system. Expanding the dataset to include more data points for comorbidities will provide a stronger foundation for these refinements. Investigating the impact of comorbidities on BCG signals could further optimize the algorithm for specific patient populations, enabling customization for individual health profiles. As the demand for contactless patient monitoring grows, ongoing updates and algorithm refinements based on real-world data, combined with enriched datasets, will be essential to maintaining the algorithm's accuracy, adaptability, and relevance in clinical practice.

Conclusion

In summary, this study presents a novel CNN-based, contactless HR measurement algorithm using BCG signals, designed for reliable, continuous HR monitoring in diverse clinical environments. By integrating advanced signal processing with ML, this approach achieves a high degree of accuracy, adaptability, and robustness, establishing a new benchmark for noninvasive monitoring solutions. The algorithm's capacity to perform reliably across varied demographics and conditions highlights its potential as a transformative tool for real-time HR monitoring, promising improved patient outcomes through timely detection and interventions. Future research will focus on expanding the algorithm's validation across various clinical environments and integrating additional vital signs to create a more holistic and effective monitoring solution.

Acknowledgments

Generative artificial intelligence tools were used in the preparation of this manuscript for tasks, such as language refinement. All content, references, and citations were verified and curated by the authors, who remain fully responsible for the accuracy, originality, and integrity of the manuscript.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Eddahchouri Y, Peelen RV, Koeneman M, Touw HRW, van Goor H, Bredie SJH. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. Br J Anaesth 2022 May;128(5):857-863. [doi: 10.1016/j.bja.2022.01.036] [Medline: 35282866]
2. Stellpflug C, Pierson L, Roloff D, et al. Continuous physiological monitoring improves patient outcomes. Am J Nurs 2021 Apr 1;121(4):40-46. [doi: 10.1097/01.NAJ.0000742504.44428.c9] [Medline: 33755624]
3. Yadav A, Dandu H, Parchani G, et al. Early detection of deteriorating patients in general wards through continuous contactless vital signs monitoring. Front Med Technol 2024;6:1436034. [doi: 10.3389/fmedt.2024.1436034] [Medline: 39328308]

XSL•FO
RenderX

4. Vincent JL, Einav S, Pearse R, et al. Improving detection of patient deterioration in the general hospital ward environment. Eur J Anaesthesiol 2018 May;35(5):325-333. [doi: 10.1097/EJA.0000000000000798] [Medline: 29474347]
5. Serhani MA, El Kassabi HT, Ismail H, Nujum Navaz A. ECG monitoring systems: review, architecture, processes, and key challenges. Sensors (Basel) 2020 Mar 24;20(6):1796. [doi: 10.3390/s20061796] [Medline: 32213969]
6. Biswas D, Simoes-Capela N, Van Hoof C, Van Helleputte N. Heart rate estimation from wrist-worn photoplethysmography: a review. IEEE Sensors J 2019 Aug 15;19(16):6560-6570. [doi: 10.1109/JSEN.2019.2914166]
7. Kumar A, Komaragiri R, Kumar M. A review on computation methods used in photoplethysmography signal analysis for heart rate estimation. Arch Computat Methods Eng 2022 Mar;29(2):921-940. [doi: 10.1007/s11831-021-09597-4]
8. Rafiei H, Amiri M, Moghaddasi J. Skin irritation because of electrocardiograph lead in patients in intensive care unit. Int Wound J 2013 Feb;10(1):116. [doi: 10.1111/j.1742-481X.2011.00915.x] [Medline: 22221961]
9. Torp KD, Modi P, Pollard EJ, Simon LV. Pulse oximetry. In: StatPearls: StatPearls Publishing; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK470348/ [accessed 2026-04-15]
10. Pinheiro E, Postolache O, Girão P. Theory and developments in an unobtrusive cardiovascular system representation: ballistocardiography. Open Biomed Eng J 2010;4(1):201-216. [doi: 10.2174/1874120701004010201] [Medline: 21673836]
11. Gubner RS, Rodstein M, Ungerleider HE. Ballistocardiography; an appraisal of technic, physiologic principles, and clinical value. Circulation 1953 Feb;7(2):268-286. [doi: 10.1161/01.cir.7.2.268] [Medline: 13019942]
12. Vogt E, MacQuarrie D, Neary JP. Using ballistocardiography to measure cardiac performance: a brief review of its history and future significance. Clin Physiol Funct Imaging 2012 Nov;32(6):415-420. [doi: 10.1111/j.1475-097X.2012.01150.x] [Medline: 23031061]
13. Tramontano A, Tamburis O, Cioce S, Venticinque S, Magliulo M. Heart rate estimation from ballistocardiogram signals processing via low-cost telemedicine architectures: a comparative performance evaluation. Front Digit Health 2023;5:1222898. [doi: 10.3389/fdgth.2023.1222898] [Medline: 37583833]
14. Ahmed N, Srivyshnav KS, Chokalingam K, et al. Classification of sleep-wake state in ballistocardiogram system based on deep learning. Presented at: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 11-15, 2022. [doi: 10.1109/EMBC48229.2022.9871831]
15. Saran V, Kumar G, Parchani G. Unsupervised extraction of respiration cycles through ballistocardiography. Presented at: Third International Conference on Advanced Informatics for Computing Research, ICAICR 2019; Jun 15-16, 2019. [doi: 10.1007/978-981-15-0108-1_14]
16. Parchani G, Kumar G, Rao R, Udupa K, Saran V. Efficacy of non-contact ballistocardiographysystem to determine heart rate variability. Ann Neurosci 2022 Jan;29(1):16-20. [doi: 10.1177/09727531211063426] [Medline: 35875429]
17. Saran V, Kumar G, Dhawan U, Parchani G. Unsupervised identification of cardiac contraction through ballistocardiography. Presented at: 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); Dec 3-6, 2018. [doi: 10.1109/IECBES.2018.8626652]
18. Saran V, Kumar R, Kumar G, Chokalingam K, Rawooth M, Parchani G. Validation of Dozee, a ballistocardiography-based device, for contactless and continuous heart rate and respiratory rate measurement. Annu Int Conf IEEE Eng Med Biol Soc 2022 Jul;2022:1939-1943. [doi: 10.1109/EMBC48229.2022.9871007] [Medline: 36086663]
19. Kedia S, Ur Rahman I, Kaushal A, et al. Clinical validation of an indigenous micro-vibration vital parameter monitor Dozee VS. Presented at: 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS); Jan 3-8, 2023. [doi: 10.1109/COMSNETS56262.2023.10041343]
20. Suliman A, Carlson C, Ade CJ, Warren S, Thompson DE. Performance comparison for ballistocardiogram peak detection methods. IEEE Access 2019;7:53945-53955. [doi: 10.1109/ACCESS.2019.2912650]
21. Jiang W, Majumder S, Kumar S, et al. A wearable tele-health system towards monitoring COVID-19 and chronic diseases. IEEE Rev Biomed Eng 2022;15:61-84. [doi: 10.1109/RBME.2021.3069815] [Medline: 33784625]
22. Yang T, Yuan H, Yang J, et al. Solving variability: accurately extracting feature components from ballistocardiograms. Digit Health 2024;10:20552076241277746. [doi: 10.1177/20552076241277746] [Medline: 39247094]
23. Javaid AQ, Toreyin H, Inan OT. Quantifying the effects of blood pressure changes on ballistocardiogram signals. Presented at: 2017 25th Signal Processing and Communications Applications Conference (SIU); May 15-18, 2017. [doi: 10.1109/SIU.2017.7960423]
24. Sadek I, Biswas J, Abdulrazak B. Ballistocardiogram signal processing: a review. Health Inf Sci Syst 2019 Dec;7(1):10. [doi: 10.1007/s13755-019-0071-7] [Medline: 31114676]
25. Mai Y, Chen Z, Yu B, Li Y, Pang Z, Han Z. Non-contact heartbeat detection based on ballistocardiogram using UNet and bidirectional long short-term memory. IEEE J Biomed Health Inform 2022 Aug;26(8):3720-3730. [doi: 10.1109/JBHI.2022.3162396] [Medline: 35333727]
26. Huang Y, Jin T, Sun C, Li X, Yang S, Zhang Z. Efficient J peak detection from ballistocardiogram using lightweight convolutional neural network. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:269-272. [doi: 10.1109/EMBC46164.2021.9630255] [Medline: 34891288]
27. Good machine learning practice for medical device development: guiding principles. : U.S. Food and Drug Administration (FDA); 2025 URL: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles [accessed 2026-05-18]

XSL•FO
RenderX

28. Vesterinen V, Rinkinen N, Nummela A. A contact-free, ballistocardiography-based monitoring system (Emfit QS) for measuring nocturnal heart rate and heart rate variability: validation study. JMIR Biomed Eng 2020;5(1):e16620. [doi: 10.2196/16620]
29. Clemente J, Valero M, Li F, Wang C, Song W. Helena: real-time contact-free monitoring of sleep activities and events around the bed. Presented at: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom); Mar 23-27, 2020. [doi: 10.1109/PerCom45495.2020.9127369]
30. Pröll SM, Tappeiner E, Hofbauer S, Kolbitsch C, Schubert R, Fritscher KD. Heart rate estimation from ballistocardiographic signals using deep learning. Physiol Meas 2021 Jul 28;42(7):075005. [doi: 10.1088/1361-6579/ac10aa] [Medline: 34198282]
31. Jung H, Kimball JP, Receveur T, Agdeppa ED, Inan OT. Accurate ballistocardiogram based heart rate estimation using an array of load cells in a hospital bed. IEEE J Biomed Health Inform 2021 Sep;25(9):3373-3383. [doi: 10.1109/JBHI.2021.3066885] [Medline: 33729962]
32. Breteler MJM, KleinJan EJ, Dohmen DAJ, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology 2020 Mar;132(3):424-439. [doi: 10.1097/ALN.0000000000003029] [Medline: 31743149]
33. Kholghi M, Szollosi I, Hollamby M, Bradford D, Zhang Q. A validation study of a ballistocardiograph sleep tracker against polysomnography. J Clin Sleep Med 2022 Apr 1;18(4):1203-1210. [doi: 10.5664/jcsm.9754] [Medline: 34705630]
34. Geng D, Yin Y, Fu Z, et al. Heart rate detection method based on ballistocardiogram signal of wearable device: algorithm development and validation. Heliyon 2024 Mar 15;10(5):e27369. [doi: 10.1016/j.heliyon.2024.e27369] [Medline: 38486774]
35. Sumali B, Mitsukura Y, Nishimura T. Contactless continuous heart rate monitoring system using ballistocardiography. PLoS ONE 2022;17(7):e0272072. [doi: 10.1371/journal.pone.0272072] [Medline: 35905114]
36. Feng J, Huang W, Jiang J, et al. Non-invasive monitoring of cardiac function through ballistocardiogram: an algorithm integrating short-time Fourier transform and ensemble empirical mode decomposition. Front Physiol 2023;14:1201722. [doi: 10.3389/fphys.2023.1201722] [Medline: 37664434]
37. Pino EJ, Chavez JAP, Aqueveque P. BCG algorithm for unobtrusive heart rate monitoring. Presented at: 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); Nov 6-8, 2017. [doi: 10.1109/HIC.2017.8227614]
38. Moukadem A, Finnaoui A, Gassara HE, Adolphe D, Schacher L, Dieterlen A. Time-frequency domain for BCG analysis. Presented at: 2018 International Conference on Computer and Applications (ICCA); Aug 25-26, 2018. [doi: 10.1109/COMAPP.2018.8460408]
39. Schranz C, Halmich C, Mayr S, Heib DPJ. Surrogate modelling of heartbeat events for improved J-peak detection in BCG using deep learning. Front Netw Physiol 2024;4:1425871. [doi: 10.3389/fnetp.2024.1425871] [Medline: 39099720]
40. Choe ST, Cho W. Simplified real-time heartbeat detection in ballistocardiography using a dispersion-maximum method. Biomed Res (Tokyo) 2017;28:3974-3985 [FREE Full text]
41. Ghosh A, Anand N, Kaushik P, Bagrodia V, Pal PK, Yadav R. IOT based sensor system for 24×7 monitoring movement disorder symptoms using machine learning. Presented at: 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS); Jan 3-8, 2023. [doi: 10.1109/COMSNETS56262.2023.10041330]
42. Lakshman P, Gopal PT, Khurdi S. Effectiveness of remote patient monitoring equipped with an early warning system in tertiary care hospital wards: retrospective cohort study. J Med Internet Res 2025 Jan 15;27:e56463. [doi: 10.2196/56463] [Medline: 39813676]

Abbreviations

BCG: ballistocardiography
CNN: convolutional neural network
DR: detection rate
ECG: electrocardiography
FDA: Food and Drug Administration
HR: heart rate
ICU: intensive care unit
LOA: limits of agreement
MAE: mean absolute error
ML: machine learning
STFT: short-time Fourier transform

XSL•FO
RenderX

Edited by KC Wong; submitted 15.Jan.2025; peer-reviewed by HW Chiu, Z Lu; revised version received 14.Nov.2025; accepted 28.Jan.2026; published 30.Jul.2026.

Please cite as:
Chokalingam K, Saravanan M, Kaushal A, Rao S, Ur Rahman I, Nambiar A, Dandwate M, Mahajan R, Sarkar K, Parchani G
A Novel Convolutional Neural Network–Based Algorithm for Heart Rate Measurement From Ballistocardiography Signals in Diverse Clinical Settings: Observational Study
JMIR Cardio 2026;10:e71224
URL: https://cardio.jmir.org/2026/1/e71224
doi:10.2196/71224

© Kumar Chokalingam, Muthukumarasamy Saravanan, Ashish Kaushal, Srishti Rao, Inam Ur Rahman, Ashwathi Nambiar, Mudit Dandwate, Ravi Mahajan, Kunal Sarkar, Gaurav Parchani. Originally published in JMIR Cardio (https://cardio.jmir.org), 30.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Deep Neural Networks for Automatic Atrial Fibrillation Detection Using Long-Term Ambulatory Electrocardiography: Retrospective Diagnostic Accuracy Study

Jagdeep Sedha[1,2*], MD; Jukka A Lipponen[3*], MSc, PhD; Antti Aho[1], MD; Helena Jäntti[4], MD, PhD; Onni E Santala[1,4], MD, PhD; Tomi P Laitinen[1,5], MD, PhD; Tiina M Laitinen[3], MSc, PhD; Jari Halonen[1,2], MD, PhD; Mika P Tarvainen[3], MSc, PhD; Eemu-Samuli Seljola[1,6,7], MD, PhD; Noora S Naukkarinen[1], MD; Olli Rantula[1,2], MD, MSc; Tuomas T Rissanen[1,8], MD, PhD; Juha E K Hartikainen[1,2], MD, PhD; Tero J Martikainen[9], MD, PhD

[1]School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 8, Kuopio, Finland

[2]Heart Center, Kuopio University Hospital, Kuopio, Finland

[3]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland

[4]Department of Anesthesiology and Intensive Care, North Karelia Central Hospital, Joensuu, Finland

[5]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland

[6]Mikkeli Central Hospital, Mikkeli, Finland

[7]University of Helsinki, Helsinki, Finland

[8]Heart Center, North Karelia Central Hospital, Joensuu, Finland

[9]Department of Prehospital Emergency Care, Acute Services, Kuopio, Finland

[*]these authors contributed equally

Corresponding Author:
Jagdeep Sedha, MD
School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 8, Kuopio, Finland

Abstract

Background: Atrial fibrillation (AF), the most prevalent cardiac arrhythmia, affects 2% to 4% of the global adult population and is associated with an increased risk of stroke. Early diagnosis of AF and atrial flutter (AFL) is crucial due to their association with stroke risk and the challenge posed by their often asymptomatic and episodic nature. Traditional electrocardiogram (ECG) interpretation requires substantial expert input and can be challenging, especially with poor-quality ECGs.

Objective: This study aimed to evaluate the performance of a deep neural network (DNN) model in detecting AF/AFL from a large, heterogeneous set of long-term ambulatory ECG recordings, including clinical data collected over 6 months at a university hospital, and assess its effectiveness in a setting reflecting the diversity and complexity of real-world clinical data.

Methods: The research combined public datasets totaling 10,248 patients, ECG data from our previous studies (648 patients), and authentic long-term ECG recordings from 4346 patients at Kuopio University Hospital for development of the DNN model. Its clinical accuracy and generalizability were assessed using a separate test dataset consisting of 1039 pseudonymized long-term ECG recordings from 1010 patients, all thoroughly reviewed and annotated by experts.

Results: The DNN model demonstrated high effectiveness, achieving 96.4% sensitivity and more than 99.99% specificity for time-level AF and AFL detection. At the recording level, it identified AF and AFL with 100% sensitivity and 98.77% specificity, producing false positives in only 1.2% (11/897) of recordings, of which 81.8% (9/11) had other non–AF/AFL arrhythmias. The model maintained high performance across diverse patient characteristics, including varying ages, comorbidities, coexisting arrhythmias, and poor-quality ECG recordings.

Conclusions: The results demonstrate that the proposed DNN model may support automated screening for AF and AFL in long-term ambulatory ECG recordings and may reduce manual review workload in clinical practice.

(JMIR Cardio 2026;10:e83714) doi:10.2196/83714

KEYWORDS

atrial fibrillation; deep neural networks; electrocardiogram; electrocardiogram analysis; ECG analysis; cardiac arrhythmia detection; artificial intelligence; AI; machine learning

XSL•FO
RenderX

Introduction

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia, affecting 2% to 4% of the global adult population [1]. AF, as well as atrial flutter (AFL), is associated with an increased risk of stroke and substantial health care costs [1]. The prevalence of AF and AFL is predicted to increase 2.3-fold by 2060 due to the aging population [1-4]. The diagnosis of AF and AFL requires electrocardiogram (ECG) documentation. However, AF and AFL are often intermittent and asymptomatic. Thus, standard 24- or 48-hour ECG recordings are often insufficient for detecting AF and AFL, making early detection and prevention challenging [5-7].

AF is characterized by disorganized atrial electrical activity, resulting in an irregular ventricular response and absence of discrete P waves on surface ECG recordings [8-11]. In contrast, AFL is a more organized atrial arrhythmia, typically driven by a macroreentrant circuit and associated with regular flutter waves [12]. Although AF and AFL share common risk factors and may coexist, their ECG manifestations differ in ways that are relevant for automated rhythm detection [9-11].

From an automated ECG analysis perspective, AFL presents a particular challenge due to variable atrioventricular (AV) conduction. Irregular ventricular responses in AFL can closely resemble those of AF on surface ECG, leading to frequent misclassification by both rule-based and deep learning–based methods [13-15]. As a result, approaches relying primarily on RR interval irregularity may struggle to distinguish AFL with variable AV conduction from AF [6-8].

AF may be classified as paroxysmal or persistent based on episode duration and termination characteristics [16,17]. Paroxysmal AF is intermittently present and, therefore, more likely to be missed in long-term ambulatory ECG recordings, whereas persistent AF is continuously detectable [16,17].

In addition to true atrial arrhythmias, several cardiac rhythm disturbances such as supraventricular ectopic beats (SVEBs), AV conduction abnormalities, and atrial tachycardias may mimic AF or AFL on surface ECG, increasing the risk of false-positive detections in automated ECG analysis [18,19].

Recently, longer ECG recordings lasting a week or even longer have become available owing to the emergence of novel ECG recording devices such as wearable single-channel ECG devices [20,21]. However, longer ECG recordings typically also encompass 10% to 30% of noisy data and artifacts [22,23] arising from motion-related interference or muscular activity [20,21]. Consequently, contemporary analysis techniques require substantial manual effort from health care personnel, particularly when analyzing noisy ECG data [20,21]. Given the growing incidence of cardiac arrhythmias and the increasing duration of recordings, there is an unmet demand for automated approaches capable of accurate detection of cardiac arrhythmias [20,21].

In contrast to the traditional stepwise method of automated arrhythmia analysis involving signal preprocessing, feature extraction, and classification [24], our study focused on the application of deep neural networks (DNNs). These networks offer comprehensive, end-to-end processing of raw ECG data [21], potentially uncovering new diagnostic features from large volumes of unprocessed data.

Several end-to-end DNN models have been proposed for ECG-based AF detection [25-34] and arrhythmia analysis [7,21]. Most of these models have been developed and evaluated using small public databases with a limited number of recordings. This limitation hampers the development of models that are generalizable, resulting in poor performance when applied to unseen data. While a few recent studies have applied DNNs to long-term ambulatory recordings, they have not assessed performance at the full recording or patient level in real-world clinical datasets [35,36].

Furthermore, only a few models have been designed and tested using a substantial amount of ambulatory data, which are often affected by various artifacts, other rhythm disorders, and noise that can mimic AF [25,37,38]. The application of these DNN models for AF screening in ambulatory, free-living conditions remains an open challenge [11]. Even the most advanced DNN methods developed with substantial amounts of ECG data [28-30,35,36] still face a significant issue: a high rate of false positives when applied to patient-operated ambulatory ECGs collected under real-world conditions.

In this study, we used an extensive set of authentic clinical data and specifically investigated the capability of a DNN model for automatic identification of patients with AF and AFL based on 24- and 48-hour ECG recordings while highlighting pertinent ECG segments exceeding 30 seconds for physician review.

Methods

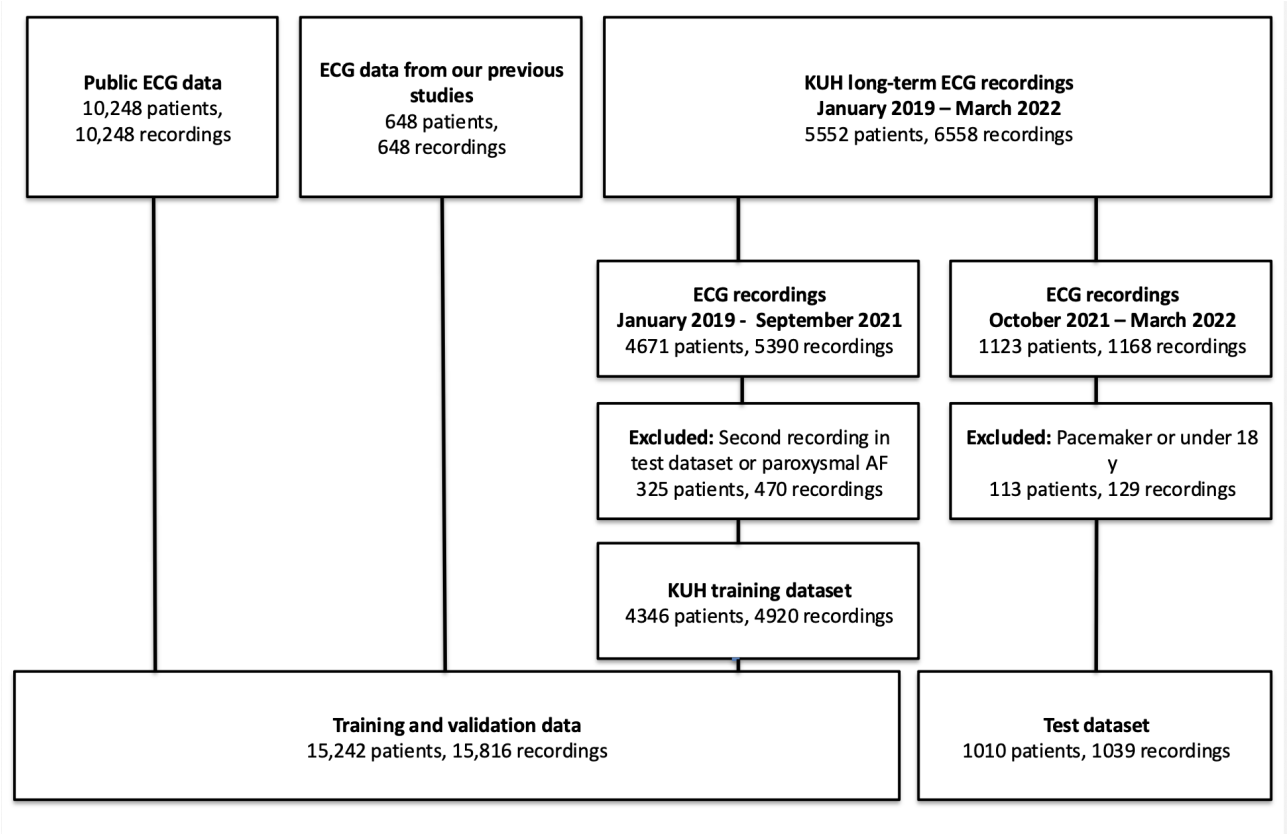
Training and Validation Data

The training and validation dataset comprised 10,896 recordings from 10,896 patients. Of these, 10,248 recordings were gathered from a public ECG dataset, while the remaining 648 recordings were obtained from our previous studies designed for training and validation of machine learning methods for arrhythmia detection (Figure 1, Table 1). The flowchart illustrating the training and validation data is presented in the Results section. Public datasets provided AF and AFL labels that were used as is, whereas ECG recordings from previous studies were annotated by trained physicians. Additionally, the training and validation dataset included samples from long-term ECG recordings at Kuopio University Hospital (KUH). The KUH training dataset encompasses recordings from January 2019 to September 2021 (4920 recordings from 4346 patients). Clinical ECG data from KUH were labeled according to standard clinical practice and analyzed using the Medilog Darwin software (Schiller). During the clinical review process, arrhythmic events, including AF and AFL, were identified and verified by experienced ECG analysis specialists. In addition, ectopic beats were classified as ventricular ectopic beats or SVEBs and verified during expert review. ECG segments in which all 3 channels were too noisy for reliable identification of QRS complexes were labeled as nonanalyzable. Recordings from patients with paroxysmal AF were excluded from the training and validation data and moved to the KUH test dataset (see below). In addition, patients who were identified in the KUH

XSL•FO
RenderX

test dataset were excluded from the KUH training and validation dataset. Across all training datasets, including public datasets, the total number of patients was 15,242, of whom 3180 (20.9%) had AF or AFL.

Figure 1. Study flowchart for electrocardiogram (ECG) recordings illustrating the division into training and validation data and a separate test dataset, with exclusions noted. Patient counts in the period-specific KUH subsets are not mutually exclusive because some patients had ECG recordings in both time periods. These patients are counted in both period-specific boxes but only once in the total KUH patient count. AF: atrial fibrillation; KUH: Kuopio University Hospital.



XSL•FO
RenderX

Table . Overview of training and validation datasets detailing the number of electrocardiogram (ECG) recordings, the number of patients per dataset, and the duration of ECG recordings for each dataset.

Database	Number of recordings (n=15,816), n (%)	Number of patients (n=15,242), n (%)	Recording duration
KUH[a] training dataset	4920 (31.1)	4346 (28.5)	24 h
ECG data from our previous studies			
Arrhythmia detection and 1-time ECG [39,40]	461 (2.9)	461 (3.0)	0.5-1 h
AFIB24h[b] dataset [41]	187 (1.2)	187 (1.2)	24 h
Public ECG data			
Long-Term AF[c] Database [42]	84 (0.5)	84 (0.6)	24 h
Chapman University dataset [43]	10,000 (63.2)	10,000 (65.6)	10 s
Smart Health for Assessing the Risk of Events via ECG dataset Database [44]	139 (0.9)	139 (0.9)	24 h
MIT[d]-BIH[e] Atrial Fibrillation Database [45]	25 (0.2)	25 (0.2)	10 h

[a]KUH: Kuopio University Hospital.

[b]AFIB24h: Atrial Fibrillation Detection: 24 Hour Study.

[c]AF: atrial fibrillation.

[d]MIT: Massachusetts Institute of Technology.

[e]BIH: Beth Israel Hospital.

Altogether, the training and validation data contained ECG samples from 15,242 patients obtained using various ECG recorders and from different geographical areas. A randomly selected 20% of the training dataset recordings were used as a validation dataset to perform DNN hyperparameter tuning. The training, validation, and test datasets had separate sets of patients. The initial split was performed at the recording level. Subsequently, all recordings from patients appearing in the test set were removed from the validation or training sets, and recordings from patients appearing in the validation set were removed from the training set. As a result, the final training, validation, and test datasets had mutually exclusive patient groups, whereas the test dataset could contain multiple recordings from the same patient. Public datasets were derived from independently collected cohorts in the United States, China, and Italy, whereas the local dataset was collected at KUH. Given the independent origins of these datasets, patient overlap between sources was not expected, thereby reducing the risk of data leakage across training, validation, and test sets.

Test Dataset

Clinical accuracy and the generalizability of the method were tested using 1039 retrospective pseudonymized long-term ECG recordings from 1010 patients (test dataset) conducted at KUH between October 2021 and March 2022. The exclusion criteria were pacemaker use (n=24, 2.3% of the recordings) and patients under 18 years (n=105, 10.1% of the recordings). Arrhythmias underwent electrophysiologist review including arrhythmia classification identified during clinical practice analyzed using the Medilog Darwin software. In addition, segments whose 3 channels were too noisy for identification of QRS complexes were assessed as nonanalyzable. ECG recordings in the test datasets underwent an additional rhythm annotation procedure where researchers verified and corrected (if needed) on a second-by-second basis all AF and AFL segments. Rhythm annotations made during clinical practice were used as a starting point. The AF class combined both AF and AFL rhythms. Medication data were obtained through automated extraction from the hospital electronic health record system at KUH.

Development of the DNN Model

We developed a convolutional DNN model specifically designed for the detection of AF and AFL. We selected a residual network as the core architecture, incorporating shortcut connections between every second convolutional layer to address the vanishing gradient problem [46,47]. A similar residual network has been successfully applied to an ECG arrhythmia classification task before [21].

The model processes a 10-second segment of raw ECG data as input, consisting of 3 channels sampled at 125 Hz, and requires no additional patient-related information. The model classifies the 10-second ECG input into 3 categories: AF or AFL, non–AF/AFL, or nonanalyzable. Segment-level classification was performed using the maximum of the output without class-specific probability thresholds. The number of convolutional filters was doubled after every 4 convolutional layers [21]. Prior to each convolutional layer, batch normalization and a rectified linear unit activation were applied [48].

The network weights were initialized using the initialization method by He et al [49]. For optimization, the Adam optimizer [50] was used with a batch size of 128 and an initial learning rate of 10^{-3}. Additionally, dropout with a probability of 0.1 was

XSL•FO
RenderX

used to prevent overfitting [51]. The model was trained using a categorical cross-entropy loss. To improve training efficiency, the learning rate was reduced by a factor of 10 if the validation loss did not improve for 2 consecutive epochs.

Model hyperparameters—including the number of convolutional layers (search space: 17 to 37), number of filters (8 to 32), filter size, and dropout rate (0 to 0.4)—were optimized using grid search. The model with the lowest validation loss during training was selected for final evaluation.

The final optimized network consists of 37 convolutional layers. The convolutional layers initially use 16 filters with a filter length of 12, and the number of filters increases after every 4 layers. A schematic overview of the DNN architecture is presented in Figure S1 in Multimedia Appendix 1.

The model output was subsequently postprocessed to account for the temporal continuity of detected rhythms. AF detections shorter than 30 seconds were excluded as clinically insignificant in accordance with current European Society of Cardiology guidelines defining clinically relevant AF episodes as lasting at least 30 seconds [2]. Closely spaced, clinically significant AF and AFL segments were merged into unified episodes using a 4-minute window to enhance detection sensitivity and provide more meaningful information for clinical interpretation.

DNN Model Evaluation

All 1039 long-term ambulatory recordings in the test dataset were segmented into fixed, nonoverlapping 10-second windows and input into the DNN model. Each segment was scaled to a fixed range between −1 and 1 to reduce amplitude variability across recordings. The DNN produced 1 rhythm classification per 10-second segment (AF or AFL, non–AF/AFL, or nonanalyzable) based on signal quality and rhythm characteristics, with no segments excluded from analysis. As ECG recordings are annotated to true classes second by second, model accuracy can be assessed at the time and recording level. Time-level accuracy of AF and AFL detection was evaluated by comparing annotations to the closest prediction and deemed to be correctly classified if the annotation and prediction matched. At the recording level, patients with AF or AFL were deemed correctly detected (true positive) if there was at least one correctly detected 30-second AF or AFL segment. Patients without AF or AFL were correctly identified (true negative) if the DNN model did not predict AF or AFL in any 30-second sequence for the patient. Evaluation at the recording level is useful for assessing how the DNN model can be applied to the analysis of online ECG recordings and notify health care personnel of whether AF or AFL is detected from the ongoing recording. The time-level evaluation is similar in clinical applications such as offline ambulatory ECG analysis, where it is critical to identify the onset and offset of an AF or AFL segment.

Ethical Considerations

This study was approved by the ethics committee of the Northern Savo Hospital District (2061/2020). This study was conducted according to the principles of the Declaration of Helsinki. No informed consent from patients was needed due to the retrospective nature of the study. All data were pseudonymized before analyses. This study used retrospective ECG data owned by KUH and was conducted as part of institutional research activities. The manuscript and supplementary materials do not include images or other content that would allow for the identification of individual participants.

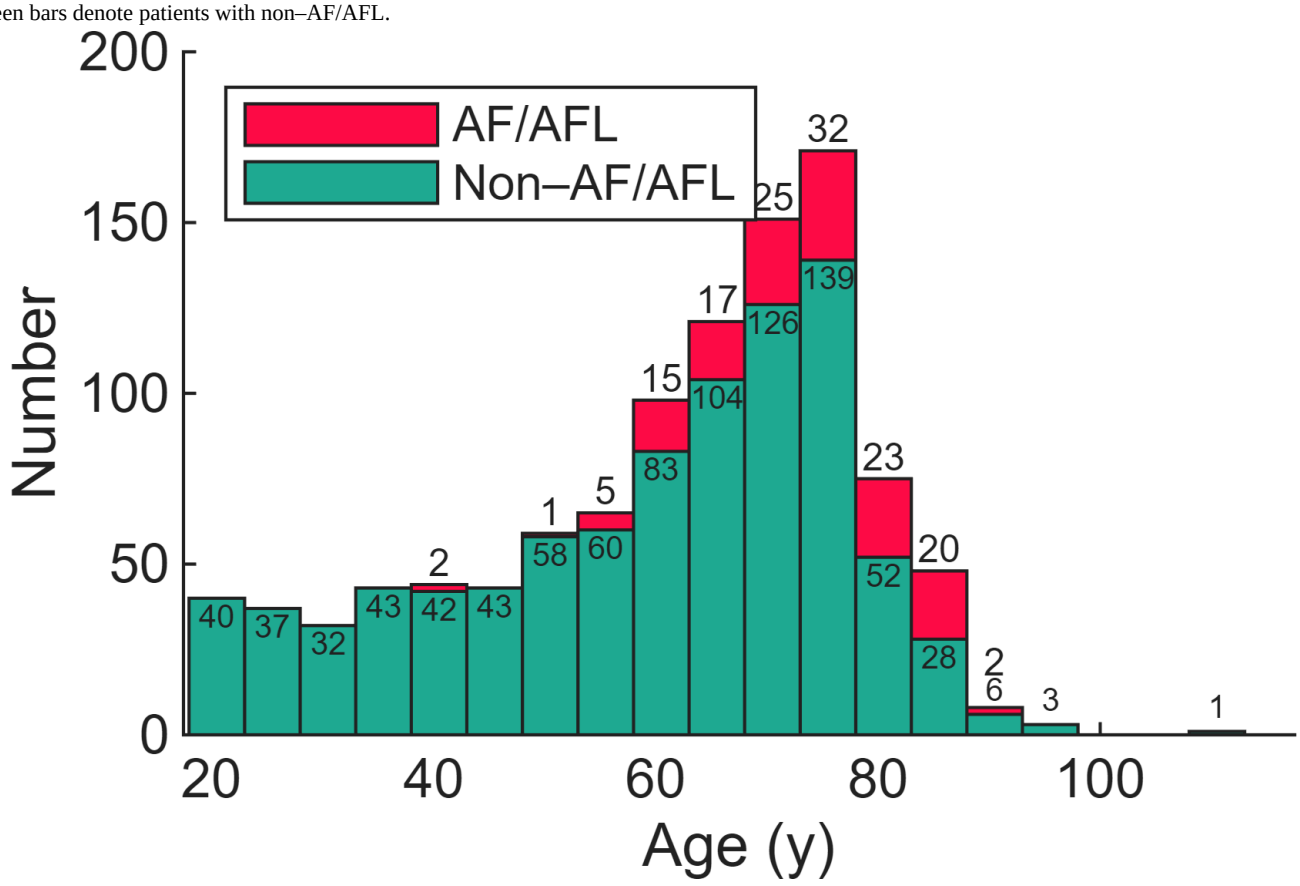
Results

Overview

The final test dataset consisted of 1039 three-channel ECG recordings ranging from 24 to 48 hours in duration from 1010 unique individuals, as shown in the Figure 1 flowchart. The mean age of the patients was 60.37 (SD 17.75; range 18-105) years, and 48% (485/1010) were men. AF or AFL was present in 13.7% (142/1039) of the ECG recordings, of which 16.2% (23/142) exhibited both AF and AFL and non–AF/AFL arrhythmias. Figure 2 shows the age distribution and rhythm distribution of the test dataset. In total, there were 25,072 hours of ECG data, with 3028.8 (12.1%) hours of AF and AFL. AF and AFL were most prevalent in individuals aged 60 to 90 years. In addition, 0.4% (92.9/25,072) of hours of data were deemed nonanalyzable.

XSL•FO
RenderX

Figure 2. Age distribution of the test set (n=1039) recordings. The red bars denote patients with atrial fibrillation (AF) or atrial flutter (AFL), and the green bars denote patients with non–AF/AFL.



Data on medication were available for 66.6% (673/1010) of the dataset patients (Table 2). Notably, oral anticoagulants were used in 51.5% (50/97) of patients in the AF and AFL group. Antihypertensive drugs were the most used medication in both the AF and AFL group and non–AF/AFL group. Non–AF/AFL rhythms are shown in Table 3. Cardiologist statements on arrhythmias were available for 88.1% (915/1039) of the recordings. Notably, combined cases of ventricular ectopic beats (>1000) and SVEBs (>1000) accounted for 28.2% (258/951) of the non–AF/AFL arrhythmias. Other significant observations included supraventricular tachycardias, which were present in 17.8% (163/915) of patients, and sinus bradycardia, which was present in 7.3% (67/915) of patients.

Table . Medication use in test dataset recordings.

	AF[a] or AFL[b] (n=97 recordings), n (%)	Non–AF/AFL (n=576 recordings), n (%)
Data available	97 (68.3)	576 (64.2)
No medication	0 (0)	38 (6.6)
Antiarrhythmic drugs	4 (4.1)	9 (1.6)
Antihypertensive drugs	73 (75.3)	417 (72.4)
Anticoagulants	50 (51.5)	59 (10.2)
Antithrombotic drugs	8 (8.2)	101 (17.5)
Dyslipidemia	34 (35.1)	205 (35.6)
Diabetes oral drugs	11 (11.3)	56 (9.7)
Insulin	0 (0)	19 (3.3)

[a]AF: atrial fibrillation.

[b]AFL: atrial flutter.

XSL•FO
RenderX

Table . Prevalence of non–atrial fibrillation or atrial flutter arrhythmias in the test dataset categorized by type and count, with corresponding percentages of the total cases. The table includes data on ventricular and supraventricular ectopic beats (SVEBs), atrioventricular blocks (AVBs), sinus bradycardia, and other arrhythmias identified in the patient population.

Finding	Recordings (N=1039), n (%)
Data available	915 (88.1)
VEB[a]—>100	341 (37.3)
VEB—>1000	167 (18.3)
VEB—>5000	71 (7.8)
VEB—>10,000	37 (4.0)
SVEB—>100	271 (29.6)
SVEB—>1000	91 (9.9)
SVEB—>5000	23 (2.5)
SVEB—>10,000	7 (0.8)
LBBB[b]	11 (1.2)
RBBB[c]	10 (1.1)
AVB I	66 (7.2)
AVB II	33 (3.6)
AVB III	0 (0)
Junctional rhythm	8 (0.9)
SVT[d]	163 (17.8)
VT[e]	49 (5.4)
Sinus bradycardia	67 (7.3)
Sinus pause	6 (0.7)

[a]VEB: ventricular ectopic beat.

[b]LBBB: left bundle branch block.

[c]RBBB: right bundle branch block.

[d]SVT: supraventricular tachycardia.

[e]VT: ventricular tachycardia.

Time-Level AF and AFL Detection

In the time-level classification (confusion matrix), the DNN model correctly classified 2918.2 hours as AF or AFL out of 3028.8 true hours of AF and AFL, resulting in 96.4% sensitivity for time-level AF and AFL detection (Table 4). The DNN misclassified 1.0 hours of non–AF/AFL data as AF or AFL and 0.1 hours of nonanalyzable ECG data as AF or AFL resulting in over 99.99% specificity, 99.96% positive predictive value (PPV), and 99.5% negative predictive value (NPV) for time-level AF and AFL detection.

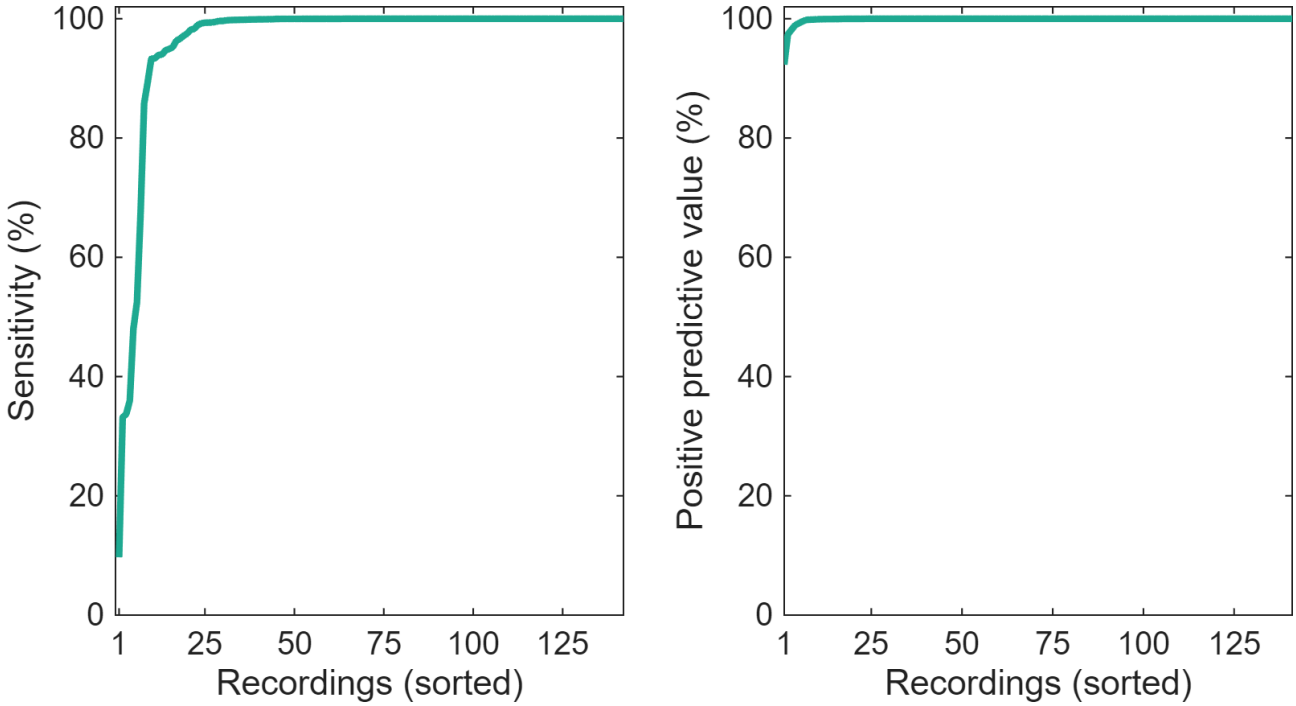
Table . Comparative analysis of true vs deep neural network (DNN)–detected hours for atrial fibrillation (AF) and atrial flutter (AFL), non–AF/AFL, and nonanalyzable categories.

	DNN-detected AF and AFL (h)	DNN-detected non–AF/AFL (h)	DNN-detected nonanalyzable (h)
True AF and AFL (h)	2918.2	109.1	1.5
True non–AF/AFL (h)	1.0	21,925.0	24.3
True nonanalyzable (h)	0.1	1.5	91.3

In the time-level analysis, the time-level sensitivity of the DNN model for detecting AF and AFL was greater than 90% in 93.7% (133/142) of AF and AFL recordings. For 2.8% (4/142) of the recordings, the sensitivity ranged between 50% and 90%, and for 3.5% (5/142) of the recordings, it was less than 50% (Figure 3). A detailed breakdown of the time-level analysis revealed that all 9 recordings with less than 90% time-level sensitivity had AFL rhythm with a 2:1, 3:1, or 4:1 (mainly 2:1) conduction ratio where part of the F waves were merged into the QRS complexes.

XSL•FO
RenderX

Figure 3. Time-level sensitivity and positive predictive value for atrial fibrillation and atrial flutter detection.



Recording-Level AF and AFL Detection

In the recording-level detection, the DNN model detected correctly at least one AF or AFL segment of more than 30 seconds in all 142 recordings containing AF or AFL. Thus, the recording-level AF and AFL detection sensitivity was 100%.

Of the 897 recordings without AF or AFL, the DNN model generated false-positive AF or AFL predictions for 11 (1.2%), with recording-level specificity of 98.77% and PPV of 92.81%. Of these 11 recordings, 9 (81.8%) showed arrhythmias: 3 contained frequent atrial premature beats, 3 contained acute rhythm changes (eg, junctional rhythm to normal SR), and 3 contained 1- or 2-degree AV block. A total of 0.2% (2/897) of the recordings showed noisy ECGs. Recording-level sensitivity, specificity, PPV, and NPV are reported in Table 5, with 95% CIs calculated using the Wilson score method.

Table . Recording-level detection performance of the deep neural network model.

Metric	Estimate (%; 95% CI)
Sensitivity	100 (97.37 - 100)
Specificity	98.77 (97.82 - 99.31)
Positive predictive value	92.81 (87.59 - 95.94)
Negative predictive value	100 (99.57 - 100)

Nonanalyzable sections of the ECG were identified in 11.1% (115/1039) of the recordings during clinical adjudication. The total duration of nonanalyzable ECGs was 92.9 hours; most of it (58.6 hours) was derived from 0.6% (6/1039) of the recordings, where lead failure or improper skin electrode contact corrupted the ECG. The DNN model correctly classified 91.3 of the 92.9 hours of nonanalyzable ECG as nonanalyzable. In addition, the DNN model labeled 25.8 hours of data as nonanalyzable that were deemed to be analyzable in the clinical workflow. A common factor between these segments was 2 channels exhibiting large ECG artifacts, whereas only 1 channel displayed visible R peaks.

Discussion

Principal Findings

This study provides a comprehensive demonstration of a deep learning approach to perform automatic AF and AFL detection from a substantial dataset encompassing more than 1000 long-term ambulatory ECG recordings from real-life clinical patients. Our DNN model successfully identified all patients exhibiting AF and AFL. Notably, it generated false AF or AFL alarms in only 1.2% (11/897) of recordings of patients who did not have AF or AFL, with 81.8% (9/11) of the false alarms due to other true arrhythmias. This demonstrates that an end-to-end DNN approach has the potential to significantly enhance the precision of algorithmic interpretation for long-term ambulatory ECG analysis. We also emphasize the importance of using a sufficiently extensive test dataset in this study, allowing for the comprehensive evaluation of the DNN model's capabilities across a diverse range of patients, including those with other arrhythmias, poor-quality ECGs, additional heart diseases, and various conditions. The training and test datasets included patients receiving a variety of rate-controlling and antiarrhythmic medications that may influence RR interval variability. The model was intentionally trained on

XSL•FO
RenderX

heterogeneous real-world data without medication-specific inputs, aiming to achieve robust AF and AFL detection independent of pharmacological modulation of heart rate dynamics.

The standard approach to automated ECG interpretation involves several steps, including preprocessing the signal; extracting features; selecting or reducing these features; and, finally, classifying them [24]. In contrast, DNNs offer a fundamentally different method by allowing a single algorithm to handle all these tasks in an end-to-end manner. This means that the DNN can process raw ECG data directly and provide diagnostic probabilities without the need for specific feature extraction [21]. If trained with sufficient raw data, DNNs could potentially learn all significant features previously identified manually as well as discover new features through a data-driven approach. Indeed, our study at the recording level showed very high sensitivity and specificity for AF and AFL detection at 100% and 98.77%, respectively. The clinical relevance of these advancements is substantial given the significant amount of unprocessed clinical data used to train the DNN algorithm. The automated detection system may not only streamline the diagnostic process but could also drastically reduce manual work, enhancing efficiency in clinical workflows. Furthermore, its capability to provide real-time alerts in online measurements with an exceedingly low rate of false alarms emphasizes its potential in real-world applications.

Comparison to Prior Work

Many prior studies examining the performance of DNN models and other methodologies for AF and AFL detection have often relied on limited ECG datasets such as the IRIDIA-AF dataset or Long-Term AF Database. These datasets typically consist of ECG recordings from a relatively small patient group [35,36,39,40]. Even DNN models developed using extensive ambulatory ECG datasets [28-30] still face a significant challenge: a high rate of false positives when applied to patient-operated ambulatory ECGs collected in real-world settings, as evidenced by a relatively low specificity of less than 99.0%. In contrast, we analyzed entire 24- to 48-hour recordings using an end-to-end DNN model that highlights clinically relevant AF and AFL segments of 30 seconds or longer for review. We evaluated its performance on a large single-center test set of 1039 recordings from real clinical patients, offering a more comprehensive and representative evaluation of DNN performance under real-world conditions. By leveraging a large and diverse training dataset collected from various ECG recorders, our model achieved comparable sensitivity while significantly improving specificity and PPV. These results highlight the potential of DNN models trained on substantial and heterogeneous datasets for improving AF detection in clinical practice.

A notable exception in long-term ambulatory ECG rhythm analysis are the studies by Zhang et al [38] and Fiorina et al [37], who evaluated DNN-based AF screening in large-scale hospital scenarios. Both DNN methods were developed using substantial ambulatory ECG datasets, yielding very promising AF screening results with a sensitivity of 99% and a specificity exceeding 98%. However, both studies focused primarily on screening patients with AF, and the capability of the DNNs to detect AF and AFL at the time level was not reported [37,38]. Importantly, our results provide strong evidence that state-of-the-art DNN models can identify patients with AF and AFL and detect AF and AFL segments in ambulatory ECGs with high temporal accuracy while effectively handling corrupted or nonanalyzable segments and maintaining a very low false positive rate regardless of the presence of other arrhythmias.

The DNN model exhibited a notably high AF and AFL time-level detection sensitivity (>90%) for 93.7% (133/142) of the patients. Upon detailed analysis, it was observed that the lower time-level sensitivity in the remaining 6.3% (9/142) of the recordings could be attributed to a specific 2:1 AFL rhythm pattern where every second P wave merged into the QRS complex. This implies that, despite its overall efficiency, the DNN model may encounter challenges in accurately identifying specific ECG patterns, as demonstrated by the 2:1 AFL rhythm. Models incorporating longer temporal context, such as convolutional neural network–bidirectional long short-term memory or transformer-based architectures, may detect transitions in AFL conduction patterns (eg, 3:1 to 2:1) and improve detection of stable 2:1 AFL segments. However, training such models would require hundreds of recordings with paroxysmal AFL and varying conduction patterns to properly capture temporal and morphological variability.

Another point of discussion is the model's performance on recordings without AF or AFL. With a very high recording-level specificity of 98.77%, the model generated false-positive AF and AFL predictions for only 1.2% (11/897) of the recordings. In fact, 81.8% (9/11) of these false-positive cases were other true arrhythmias such as frequent ectopic beats, acute rhythm changes, and 1- or 2-degree AV block. It is also notable that only a very small percentage of these conditions caused false-positive predictions. Our model also efficiently identified nonanalyzable sections of the ECG, defined as segments in which QRS complexes could not be reliably visually verified during clinical review. No separate quantitative noise index was used. Initially, the DNN model classified only 0.1% (25.8/25072.0) of hours of data as nonanalyzable, which were marked to be analyzable in the clinical workflow. False-positive nonanalyzable labeling exceeding 10 minutes occurred in 37 of 1039 recordings. Further in-depth analysis uncovered that a significant portion of these data, totaling 10 hours, originated from seven 24-hour recordings where 2 out of 3 ECG channels manifested significant noise or lead failure. Regarding clinical consequences, AF and AFL masking due to nonanalyzable labeling was limited. As shown in the confusion matrix, approximately 0.1% (1.5/3028.8) of hours of true AF and AFL detections were classified as nonanalyzable, whereas a markedly larger proportion of true AF and AFL detections (109.1/3028.8, 3.6%) was misclassified as non–AF/AFL.

In the test dataset, the prevalence of AF and AFL was 13.7% (142/1039), which is lower than that reported in some real-world Holter monitoring populations [52] but substantially higher than that reported in opportunistic screening settings, where prevalence may be as low as 1% [53]. At this prevalence level, the model achieved a PPV of 92.81% (95% CI

XSL•FO
RenderX

87.59% - 95.94%) and an NPV of 100% (95% CI 99.57% - 100%; Table 5). As predictive values are inherently prevalence dependent, the high NPV observed in this study reflects the moderate prevalence of AF and AFL in the test set and indicates strong ability to rule out arrhythmia in this context. In lower-prevalence screening populations, PPV would be expected to decrease despite unchanged sensitivity and specificity, underscoring the importance of population-specific evaluation.

Subgroup analyses by age, sex, noise burden, and non-AF arrhythmia burden are presented in Table S1 in Multimedia Appendix 1. These analyses showed that AF and AFL detection sensitivity was lowest in recordings with a lower burden of non-AF arrhythmias, whereas both sensitivity and PPV were lowest in recordings with a high burden of non-AF arrhythmias. This was expected as other atrial and ventricular arrhythmias represent the most challenging confounders for differentiating AF and AFL in long-term ambulatory ECG recordings.

Methodological Considerations and Limitations

Despite the promising results, our study had certain limitations. First, we excluded patients with pacemakers from the KUH dataset (Figure 1). The pacemaker inherently maintains a steady rhythm during AF and AFL. This stabilization renders the recognition of AF and AFL from a surface ECG exceedingly challenging. Pacemaker-related artifacts are a common source of false-positive AF and AFL detections in clinical practice, and their exclusion may lead to higher specificity than in a true all-comer population. Therefore, the reported performance metrics should be interpreted in the context of this limitation.

Second, the test dataset was derived from a single center (KUH), resulting in limited ethnic diversity and limited variation in ECG devices. Although the training and validation datasets included ECG recordings from multiple systems (including Schiller, Bittium Faros, GE HealthCare MUSE 12-lead ECG, and other ambulatory 3- and 4-lead recorders), the model may still be sensitive to device-specific artifacts, which could constrain external validity and generalizability across different clinical settings, recording systems, and populations; therefore, future multicenter evaluations are warranted.

Third, patients with paroxysmal AF were excluded from the KUH training and validation datasets. The model was primarily trained on persistent AF segments, which may differ to some extent from paroxysmal AF episodes. However, because the model operates on independent 10-second ECG segments without temporal context, AF and AFL detection is based on segment-level rhythm characteristics, which limits the impact of this difference.

Fourth, the model demonstrated reduced time-level sensitivity in AFL with fixed 2:1 AV conduction ratio, which represents a clinically relevant limitation. In this rhythm pattern, atrial activity may be partially obscured by the QRS complexes, making reliable discrimination from surface ECG challenging. Consequently, this specific AFL subtype constitutes a potential blind spot for automated screening.

Fifth, from a clinical safety perspective, the proposed DNN model should be regarded as a screening and decision support tool rather than a stand-alone diagnostic system. This is particularly important given the limited interpretability of the DNN models as learned features cannot be directly linked to established physiological markers; therefore, negative automated findings should be interpreted in conjunction with clinical judgment, especially when clinical suspicion of arrhythmia remains high.

Sixth, the 4-minute merge window used to ensure continuity of longer AF and AFL segments may lead to overestimation of AF burden, particularly when multiple short episodes occur close in time.

Conclusions

The results of this study affirm the potential of using a DNN model for AF and AFL diagnostics in ambulatory settings with high sensitivity and specificity. When trained using an ample amount of data, the DNN model demonstrates the ability to distinguish ambulatory ECG changes caused by motion artifacts from AF and AFL with a low false positive rate. The presented automated detection system may streamline the diagnostic process and clinical workflows with a low false alarm rate, supporting its clinical applicability.

Acknowledgments

Generative artificial intelligence tools were used for language editing and spelling correction. No generative artificial intelligence tools were used for scientific content generation or data analysis and interpretation, and all scientific responsibility remains with the authors.

Funding

This work was supported by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (projects 5101137 and 507T044; Kuopio, Finland). JS and JAL received research support from the Finland State Research Fund. JS received support from the University of Eastern Finland (Matti and Vappu Mauko foundation).

Data Availability

The data underlying this article are not publicly available due to the privacy of the individuals who participated in the study and applicable data protection regulations. The data supporting the findings of this study will be shared on reasonable request to the corresponding author. No synthetic data were generated in this study. The code and trained model weights are not publicly available.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: JAL, JS, TM, HJ, JH, OES, ESS, TTR, MPT, TPL, TML
Data curation: JS, JAL, AA
Formal analysis: JAL, JS
Investigation: JS, JAL
Methodology: JAL
Software: JAL
Supervision: JAL, JEKH, TJM, OES
Writing—original draft: JS, JAL
Writing—review and editing: JS, JAL, AA, HJ, OES, TPL, TML, JH, MPT, ESS, NSN, OR, TTR, JEKH, TJM

Conflicts of Interest

JAL, TTR, TJM, HJ, JH, OES, ESS, and MPT are shareholders of a company (Heart2Save) that designs electrocardiogram-based software for medical equipment. JAL, MPT, OES, and HJ report personal fees from Heart2Save. The retrospective electrocardiogram data used in this study are owned by Kuopio University Hospital. The deep learning model was developed at Kuopio University Hospital as part of institutional research activities, and no data sharing or intellectual property agreements related to this study were in place with Heart2Save at the time of model development. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Deep neural network (DNN) model architecture, performance evaluation before temporal postprocessing, and analyses across demographic and signal-quality subgroups.
[PDF File, 280 KB - cardio_v10i1e83714_app1.pdf]

References

1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation 2014 Feb 25;129(8):837-847. [doi: 10.1161/CIRCULATIONAHA.113.005119] [Medline: 24345399]
2. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 1;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
3. Alberts MJ, Eikelboom JW, Hankey GJ. Antithrombotic therapy for stroke prevention in non-valvular atrial fibrillation. Lancet Neurol 2012 Dec;11(12):1066-1081. [doi: 10.1016/S1474-4422(12)70258-2] [Medline: 23153406]
4. Ruff CT, Giugliano RP, Braunwald E, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. Lancet 2014 Mar 15;383(9921):955-962. [doi: 10.1016/S0140-6736(13)62343-0] [Medline: 24315724]
5. Zhu H, Cheng C, Yin H, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. Lancet Digit Health 2020 Jul;2(7):e348-e357. [doi: 10.1016/S2589-7500(20)30107-2] [Medline: 33328094]
6. Steinberg JS, Varma N, Cygankiewicz I, et al. 2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry. Heart Rhythm 2017 Jul;14(7):e55-e96. [doi: 10.1016/j.hrthm.2017.03.038] [Medline: 28495301]
7. Kirchhof P, Bax J, Blomstrom-Lundquist C, et al. Early and comprehensive management of atrial fibrillation: executive summary of the proceedings from the 2nd AFNET-EHRA consensus conference 'research perspectives in AF'. Eur Heart J 2009 Dec;30(24):2969-2977c. [doi: 10.1093/eurheartj/ehp235] [Medline: 19535417]
8. Wigginton JG, Agarwal S, Bartos JA, et al. Part 9: adult advanced life support: 2025 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2025 Oct 21;152(16_suppl_2):S538-S577. [doi: 10.1161/CIR.0000000000001376] [Medline: 41122884]
9. Fuster V, Rydén LE, Cannom DS, et al. 2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in partnership with the European Society of Cardiology and in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. J Am Coll Cardiol 2011 Mar 15;57(11):e101-e198. [doi: 10.1016/j.jacc.2010.09.013] [Medline: 21392637]
10. Correction to: 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2024 Jun 11;149(24):e1413. [doi: 10.1161/CIR.0000000000001263] [Medline: 38857333]
11. Ko D, Chung MK, Evans PT, Benjamin EJ, Helm RH. Atrial fibrillation: a review. JAMA 2025 Jan 28;333(4):329-342. [doi: 10.1001/jama.2024.22451] [Medline: 39680399]

XSL•FO
RenderX

12. Johner N, Namdar M, Shah DC. Typical atrial flutter: a practical review. J Cardiovasc Electrophysiol 2025 Oct;36(10):2677-2691. [doi: 10.1111/jce.16542] [Medline: 39764740]
13. Rogalewski A, Plümer J, Feldmann T, et al. Detection of atrial fibrillation on stroke units: comparison of manual versus automatic analysis of continuous telemetry. Cerebrovasc Dis 2020;49(6):647-655. [doi: 10.1159/000511563] [Medline: 33207338]
14. Avidan Y, Aker A, Sliman H, Weizman B, Danon A, Kassem S. Recognizing atrial flutter in the emergency department: challenges in diagnosis. Am J Emerg Med 2025 Oct;96:224-229. [doi: 10.1016/j.ajem.2025.07.013] [Medline: 40633488]
15. Domazetoski V, Gligoric G, Marinkovic M, et al. The influence of atrial flutter in automated detection of atrial arrhythmias - are we ready to go into clinical practice? Comput Methods Programs Biomed 2022 Jun;221:106901. [doi: 10.1016/j.cmpb.2022.106901] [Medline: 35636359]
16. Michaud GF, Stevenson WG. Atrial fibrillation. N Engl J Med 2021 Jan 28;384(4):353-361. [doi: 10.1056/NEJMcp2023658] [Medline: 33503344]
17. Prystowsky EN, Padanilam BJ, Fogel RI. Treatment of atrial fibrillation. JAMA 2015 Jul 21;314(3):278-288. [doi: 10.1001/jama.2015.7505] [Medline: 26197188]
18. Chaumont C, Suffee N, Gandjbakhch E, Balse E, Anselme F, Hatem SN. Epicardial origin of cardiac arrhythmias: clinical evidences and pathophysiology. Cardiovasc Res 2022 Jun 22;118(7):1693-1702. [doi: 10.1093/cvr/cvab213] [Medline: 34152392]
19. van der Does LJ, Kharbanda RK, Teuwen CP, et al. Atrial ectopy increases asynchronous activation of the endo- and epicardium at the right atrium. J Clin Med 2020 Feb 18;9(2):558. [doi: 10.3390/jcm9020558] [Medline: 32085601]
20. Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: a review. J Electrocardiol 2019;57:S70-S74. [doi: 10.1016/j.jelectrocard.2019.08.004]
21. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69. [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
22. Holgado-Cuadrado R, Plaza-Seco C, Lovisolo L, Blanco-Velasco M. Characterization of noise in long-term ECG monitoring with machine learning based on clinical criteria. Med Biol Eng Comput 2023 Sep;61(9):2227-2240. [doi: 10.1007/s11517-023-02802-5] [Medline: 37010711]
23. Everss-Villalba E, Melgarejo-Meseguer FM, Blanco-Velasco M, et al. Noise maps for quantitative and clinical severity towards long-term ECG monitoring. Sensors (Basel) 2017 Oct 25;17(11):2448. [doi: 10.3390/s17112448] [Medline: 29068362]
24. Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J R Soc Interface 2018 Jan;15(138):20170821. [doi: 10.1098/rsif.2017.0821] [Medline: 29321268]
25. Kumar D, Peimankar A, Sharma K, Domínguez H, Puthusserypady S, Bardram JE. DeepAware: a hybrid deep learning and context-aware heuristics-based model for atrial fibrillation detection. Comput Methods Programs Biomed 2022 Jun;221:106899. [doi: 10.1016/j.cmpb.2022.106899] [Medline: 35640394]
26. Faust O, Shenfield A, Kareem M, San TR, Fujita H, Acharya UR. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. Comput Biol Med 2018 Nov 1;102:327-335. [doi: 10.1016/j.compbiomed.2018.07.001] [Medline: 30031535]
27. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Expert Syst Appl 2020 Sep;7:100033. [doi: 10.1016/j.eswax.2020.100033]
28. Ben-Moshe N, Tsutsui K, Brimer SB, Zvuloni E, Sornmo L, Behar JA. RawECGNet: deep learning generalization for atrial fibrillation detection from the raw ECG. IEEE J Biomed Health Inform 2024 Sep;28(9):5180-5188. [doi: 10.1109/JBHI.2024.3404877] [Medline: 38787663]
29. Chocron A, Oster J, Biton S, et al. Remote atrial fibrillation burden estimation using deep recurrent neural network. IEEE Trans Biomed Eng 2021 Aug;68(8):2447-2455. [doi: 10.1109/TBME.2020.3042646] [Medline: 33275575]
30. Tutuko B, Nurmaini S, Tondas AE, et al. AFibNet: an implementation of atrial fibrillation detection with convolutional neural network. BMC Med Inform Decis Mak 2021 Jul 14;21(1):216. [doi: 10.1186/s12911-021-01571-1] [Medline: 34261486]
31. Chen X, Cheng Z, Wang S, et al. Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ECG signals. Comput Methods Programs Biomed 2021 Apr;202:106009. [doi: 10.1016/j.cmpb.2021.106009] [Medline: 33631641]
32. Fan X, Yao Q, Cai Y, Miao F, Sun F, Li Y. Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings. IEEE J Biomed Health Inform 2018 Nov;22(6):1744-1753. [doi: 10.1109/JBHI.2018.2858789] [Medline: 30106699]
33. Marinucci D, Sbrollini A, Marcantoni I, Morettini M, Swenne CA, Burattini L. Artificial neural network for atrial fibrillation identification in portable devices. Sensors (Basel) 2020 Jun 24;20(12):3570. [doi: 10.3390/s20123570] [Medline: 32599796]

XSL•FO
RenderX

34. Tzou HA, Lin SF, Chen PS. Paroxysmal atrial fibrillation prediction based on morphological variant P-wave analysis with wideband ECG and deep learning. Comput Methods Programs Biomed 2021 Nov;211:106396. [doi: 10.1016/j.cmpb.2021.106396] [Medline: 34592687]
35. Krasteva V, Stoyanov T, Naydenov S, Schmid R, Jekova I. Detection of atrial fibrillation in Holter ECG recordings by ECHOView images: a deep transfer learning study. Diagnostics (Basel) 2025 Mar 28;15(7):865. [doi: 10.3390/diagnostics15070865] [Medline: 40218215]
36. Rahman MM, Rivolta MW, Vaglio M, Maison-Blanche P, Badilini F, Sassi R. Residual-attention deep learning model for atrial fibrillation detection from Holter recordings. J Electrocardiol 2025;89:153876. [doi: 10.1016/j.jelectrocard.2025.153876] [Medline: 39827741]
37. Fiorina L, Maupain C, Gardella C, et al. Evaluation of an ambulatory ECG analysis platform using deep neural networks in routine clinical practice. J Am Heart Assoc 2022 Sep 20;11(18):e026196. [doi: 10.1161/JAHA.122.026196] [Medline: 36073638]
38. Zhang P, Lin F, Ma F, et al. Automatic screening of patients with atrial fibrillation from 24-h Holter recording using deep learning. Eur Heart J Digit Health 2023;4(3):216-224. [doi: 10.1093/ehjdh/ztad018] [Medline: 37265871]
39. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 7;394(10201):861-867. [doi: 10.1016/S0140-6736(19)31721-0] [Medline: 31378392]
40. Santala OE, Lipponen JA, Jäntti H, et al. Necklace-embedded electrocardiogram for the detection and diagnosis of atrial fibrillation. Clin Cardiol 2021 May;44(5):620-626. [doi: 10.1002/clc.23580] [Medline: 33629410]
41. Santala OE, Halonen J, Martikainen S, et al. Automatic mobile health arrhythmia monitoring for the detection of atrial fibrillation: prospective feasibility, accuracy, and user experience study. JMIR Mhealth Uhealth 2021 Oct 22;9(10):e29933. [doi: 10.2196/29933] [Medline: 34677135]
42. Petrutiu S, Sahakian AV, Swiryn S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. Europace 2007 Jul;9(7):466-470. [doi: 10.1093/europace/eum096] [Medline: 17540663]
43. Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Sci Data 2020 Feb 12;7(1):48. [doi: 10.1038/s41597-020-0386-x] [Medline: 32051412]
44. Melillo P, Izzo R, Orrico A, et al. Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PLoS One 2015;10(3):e0118504. [doi: 10.1371/journal.pone.0118504] [Medline: 25793605]
45. Moody G. A new method for detecting atrial fibrillation using R-R intervals. Proc Comput Cardiol 1983;10:227-230. [doi: 10.3390/jcm12020687]
46. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision ECCV 2016: Springer; 2016:630-645. [doi: 10.1007/978-3-319-46493-0_38]
47. Xiong Z, Nash MP, Cheng E, Fedorov VV, Stiles MK, Zhao J. ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. Physiol Meas 2018 Sep 24;39(9):094006. [doi: 10.1088/1361-6579/aad9ed] [Medline: 30102248]
48. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning 2015:448-456 [FREE Full text]
49. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 IEEE International Conference on Computer Vision (ICCV) 2015:1026-1034 [FREE Full text]
50. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv. Preprint posted online on Dec 22, 2014. [doi: 10.48550/arXiv.1412.6980]
51. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929-1958 [FREE Full text] [doi: 10.5555/2627435.2670313]
52. Wu M, Lu Y, Yang W, Wong SY. A study on arrhythmia via ECG signal classification using the convolutional neural network. Front Comput Neurosci 2021;14:564015. [doi: 10.3389/fncom.2020.564015] [Medline: 33469423]
53. Lowres N, Olivier J, Chao TF, et al. Estimated stroke risk, yield, and number needed to screen for atrial fibrillation detected through single time screening: a multicountry patient-level meta-analysis of 141,220 screened individuals. PLoS Med 2019 Sep;16(9):e1002903. [doi: 10.1371/journal.pmed.1002903] [Medline: 31553733]

Abbreviations

AF: atrial fibrillation
AFL: atrial flutter
AV: atrioventricular
DNN: deep neural network
ECG: electrocardiogram
KUH: Kuopio University Hospital
NPV: negative predictive value
PPV: positive predictive value

XSL•FO
RenderX

SVEB: supraventricular ectopic beat

Edited by A Coristine; submitted 13.Sep.2025; peer-reviewed by H Mautong, S Mohanadas; revised version received 12.Apr.2026; accepted 13.Apr.2026; published 30.Jun.2026.

Please cite as:
Sedha J, Lipponen JA, Aho A, Jäntti H, Santala OE, Laitinen TP, Laitinen TM, Halonen J, Tarvainen MP, Seljola ES, Naukkarinen NS, Rantula O, Rissanen TT, Hartikainen JEK, Martikainen TJ
Deep Neural Networks for Automatic Atrial Fibrillation Detection Using Long-Term Ambulatory Electrocardiography: Retrospective Diagnostic Accuracy Study
JMIR Cardio 2026;10:e83714
URL: https://cardio.jmir.org/2026/1/e83714
doi:10.2196/83714

© Jagdeep Sedha, Jukka A Lipponen, Antti Aho, Helena Jäntti, Onni E Santala, Tomi P Laitinen, Tiina M Laitinen, Jari Halonen, Mika P Tarvainen, Eemu-Samuli Seljola, Noora S Naukkarinen, Olli Rantula, Tuomas T Rissanen, Juha E K Hartikainen, Tero J Martikainen. Originally published in JMIR Cardio (https://cardio.jmir.org), 30.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Perceptions of a Digitally Enabled Community Health Worker Intervention: Qualitative Study Among Pilot Trial Participants

Jocelyn Carter[1], MD, MPH; Natalia Swack[1], BS; Narmeen Rehman[2], MD, MPH; Yadira Reyes-Richards[1], BA; Karen Donelan[3], SCD, EdM; Anne Thorndike[1], MD, MPH

[1]Division of General Internal Medicine, Massachusetts General Hospital, 55 Fruit Street, Blake 15, Boston, MA, United States

[2]Department of Internal Medicine, Henry Ford Hospital, Detroit, MI, United States

[3]Department of Medicine, Brandeis University, 415 South Street, Waltham, MA, United States

Corresponding Author:
Jocelyn Carter, MD, MPH
Division of General Internal Medicine, Massachusetts General Hospital, 55 Fruit Street, Blake 15, Boston, MA, United States

Abstract

Background: Most studies assessing digital interventions for people with heart failure (HF) focus on clinical outcomes, and few include patient perspectives. Understanding patient experiences of the use of a digital HF platform along with community health worker (CHW) care as part of a digitally enabled CHW intervention can inform management of HF at home and improve the postdischarge phase of care.

Objective: This study aimed to identify patient perceptions related to the use of a digitally enabled CHW intervention.

Methods: This qualitative study included interviews with adults (aged ≥18 years) with HF who were assigned to the intervention arm of a pilot randomized controlled trial from September 2022 to June 2023. For 30 days after hospital discharge, intervention participants were paired with a CHW and instructed to use a digital platform that tracked biometrics (eg, heart rate, oxygenation, blood pressure, body weight, steps taken, and symptoms) and offered educational videos. In-depth interviews were conducted after the 30-day intervention was complete (between 31 and 45 days after hospital discharge). Key interview domains included barriers and facilitators to the intervention, use of remote monitoring in HF, and the role of CHWs in HF home care.

Results: Interviews with participants (N=19; mean age 62.1, SD 15.1 years) yielded five key themes: (1) the combined intervention was well received, and CHWs made the use of the digital platform more approachable; (2) the digital platform enhanced HF knowledge and confidence in self-care; (3) digital platform use was easy to integrate into daily routines; (4) in addition to assisting with navigation of unmet social needs (eg, transportation, insurance benefits, and food access), CHWs provided emotional support and increased motivation for clinical care plan adherence and platform use; and (5) connectivity issues and other technical challenges occurred with digital platform use.

Conclusions: The digital platform was easily integrated into patients' daily routines. CHWs played a key role in making the platform more approachable for participant use. Further research is needed to better understand the impact of this intervention in larger HF populations over more extended time intervals.

International Registered Report Identifier (IRRID): RR2-10.2196/55687

(JMIR Cardio 2026;10:e93288) doi:10.2196/93288

KEYWORDS

heart failure; digital platform; remote monitoring; home-based care; community health worker; social needs care

Introduction

Heart failure (HF) remains a leading cause of hospitalizations and readmissions in the United States and contributes to nearly 6.5 million inpatient hospital stays each year [1-3]. Despite policy changes, comprehensive case management strategies, and breakthroughs in pharmacological therapy [4-6], HF-related hospitalizations and their expenditures continue to rise [7]. This is, in part, driven by patients managing HF at home who often face numerous clinical (eg, tight dietary restrictions, rigid medication regimens, and daily symptom monitoring) and social (eg, transportation, food security, housing stability, and access to care) challenges [8-10]. These barriers are associated with worsening HF morbidity, mortality, and adverse outcomes [11,12]. Ultimately, these conditions can lead to HF exacerbations, increased use of urgent or emergent care, and missed opportunities for primary and preventive care after hospital discharge [13].

XSL•FO
RenderX

Digital interventions, particularly in the postdischarge setting, have emerged as a promising tool to improve HF clinical outcomes [14,15]. Digital platforms that closely monitor vital signs, activity, and symptom data have helped improve the value of remote monitoring in managing HF at home [15,16]. Most digital studies in HF are focused on usability or efficacy relevant to clinical outcomes such as reducing hospitalizations and emergency department visits [15]. Few clinical trials assessing digital interventions in HF evaluate patient perspectives on their impressions of the digital intervention itself. Understanding the patient experience with digital interventions can inform key barriers and facilitators to managing HF care at home [17,18].

In a pilot randomized controlled trial [19], a digitally enabled community health worker (CHW) intervention for patients with HF demonstrated intervention feasibility and improvements in managing HF at home. In the full sample (control: n=31; intervention: n=25), fewer participants in the intervention group were readmitted 30 days after hospital discharge compared to the control group (n=3, 12% vs n=8, 26%; *P*=.12). Both arms had similar rates of missed clinic appointments and emergency department visits. To understand the patient experience with combining a digital platform with social support from a CHW [20], qualitative interviews were performed with intervention participants, who were paired with a CHW and instructed to use a digital platform. Our study aim was to better understand the barriers and facilitators associated with use of the digitally enabled CHW intervention, CHW home care, remote monitoring, and perceptions of digital health tools in a HF population.

Methods

Setting and Study Design

This was a qualitative study of participants who received the 30-day digitally enabled CHW intervention in a pilot randomized controlled trial. All participants had a diagnosis of HF listed on their problem list within the electronic medical record, were hospitalized at the time of enrollment, and had been hospitalized at least once in the 12 months prior to enrollment. Participants were initially identified for trial participation through inpatient electronic health records at an academic medical center. Intervention participants were paired with a CHW and the digital platform. The digital platform was connected to a mobile phone app with features including a digital weight scale, a digital blood pressure monitor, a symptom questionnaire, educational videos, and an arm-worn sensor monitoring oxygenation, steps taken daily, and heart rate. The pilot trial details, including procedural flow, inclusion criteria, and study design, are described in a previous publication [19]. Of the 19 participants who completed the intervention, 19 were successfully contacted for qualitative interviews.

Interview Protocol and Measures

All interviews were conducted after the intervention via telephone 31 to 45 days after hospital discharge, and interviews lasted 20 to 45 minutes in length. An interview guide was developed through key informant interviews with patients, HF specialists, qualitative research experts, senior scientists, health services researchers, and primary care physicians. This was coupled with a review of literature associated with prior qualitative HF studies [21]. The guide covered prespecified domains, and consistency was ensured by having all interviews conducted by a senior clinical researcher (NS). None of the researchers or coauthors in this study were involved in direct patient clinical care or the delivery of the trial intervention. Pretesting of the interview guide was performed with 3 participants; no additional modifications were made and the 3 pretesting participants were included in the full sample. The interview guide included 21 items in total, and major domains were barriers and facilitators to managing HF care at home, use of CHW care in HF at home, remote monitoring in HF, and perceptions related to use of technology in HF (Multimedia Appendix 1). All procedures were approved by the Mass General Brigham Institutional Review Board (protocol number 2018P002014) on September 9, 2022, and all participants provided written informed consent.

Analysis

Transcripts from interviews were analyzed verbatim with the Dedoose (version 8.3.47b) platform. An analytic framework was developed based on the recurring themes that emerged during interviews. To optimize reliability, 2 members of the study team (NS and JC) independently reviewed the data, identified themes, and collaboratively applied the framework to categorize and interpret findings. Weekly meetings occurred for discussions focused on data interpretation and common themes. Content analysis was completed by each coder for each patient interview prior to meeting; this facilitated consensus for all generated codes. Intercoder reliability was achieved and any discrepancies were resolved with a third researcher (KD) with expertise in qualitative research and patient interviews. The available cohort was exhausted and all responses generated by interview participant responses were analyzed. Structured chart reviews were also performed via the electronic health record and stored in a Research Electronic Data Capture (REDCap; Vanderbilt University) database [22]. Data related to demographic information, insurance status, education, and major medical and psychiatric comorbidities were collected in the chart review.

Rigor

The research team used a formal report generated by the electronic medical record to identify eligible patients. Screening for eligibility was completed by research interns trained to assess eligibility, approach, and enroll participants as a part of standardized research procedures. Participants were not randomized to the intervention or control arm until all standardized study enrollment procedures were complete. Research interns were unaware of the randomization assignments until after all enrollment procedures were complete. In addition, the qualitative interview guide was developed to standardize questions asked of participants, probe any anticipated answers, and accurately capture the participant perspective. The research team was also mindful of the extent to which data can change over time during analysis and tracked all coding decisions meticulously. While the perspectives gathered from participants may not apply to those in other or non-HF populations, the themes may apply to other cohorts

XSL•FO
RenderX

with similar characteristics. The small sample size facilitated in-depth and authentic interviews reflective of the patient experience. While the nature of qualitative studies can make replication challenging, the research focus was on gathering individual perspectives to add to the evidence base for HF care at home.

Ethical Considerations

Institutional review board approval was obtained from the Mass General Brigham Human Research Committee on September 9, 2022 (2018P002014). All enrolled participants provided written informed consent prior to this study. Data were anonymized before analysis to ensure confidentiality. All study data reported were deidentified. For remuneration, US $50 was provided to participants at the time of enrollment and an additional US $200 was provided after successful study completion. All methods were carried out in accordance with guidelines and regulations outlined by the Mass General Brigham Institutional Review Board.

Results

Of the 19 pilot trial intervention participants, 19 were interviewed (Multimedia Appendix 2) after completing the 30-day digitally enabled CHW intervention. Participant characteristics included a mean age of 62.1 (SD 15.1) years, with most participants identifying as male (n=10, 53%), White (n=11, 58%), and used Medicare (n=8, 42%) or commercial insurance (n=9, 47%). Thematic analysis identified 5 key themes, highlighting how participants experienced the integrated digital platform and CHW support. These themes are described below.

Theme 1: The Combined Intervention Was Well Received, and CHWs Made the Use of the Digital Platform More Approachable

One of the most common themes described by participants was how well the intervention worked in terms of helping them manage their HF at home. Several individuals emphasized that the platform and CHW outreach complemented one another by helping reduce anxiety about self- management and facilitate engagement in care. One participant noted the following:

> *I never felt overwhelmed by the platform or working with the CHW. Just the opposite. I felt that both eased a lot of anxieties I have and helped me to filter my way through the maze.*

Participants also noted that CHW staff played a key role in making the digital platform more approachable, especially for those unfamiliar with the technology. This was particularly true for those that had not used digital platforms in the past.

Theme 2: The Digital Platform Enhanced HF Knowledge and Confidence in Self-Care

Participants reported that the intervention helped them better understand their HF and how their self-care behaviors directly impacted their health. Several individuals described gaining new knowledge about the importance of tracking vital signs, medication adherence, and recognizing early signs of exacerbation. As one participant noted,

> *This was really helpful to have the blood pressure and weight and a person that you could stay on track with. Otherwise, sometimes you could end up in the ED for something small if you didn't have clear information or hear back from your PCP quickly enough.*

Others described how education from the intervention and real-time feedback from the monitoring tools prompted changes in their behavior, including more consistent medication use and routine self-monitoring.

Theme 3: Digital Platform Use Was Easy to Integrate Into Daily Routines

Participants reported that the digital health platform fit easily into their daily routines and supported consistent engagement with self-care activities. Several individuals described the process of using the device as quick and manageable, with one noting,

> *My daily sessions with the technology were less than 10 minutes in the morning to do the blood pressure and weight which was great.*

Others emphasized how the platform helped reinforce new habits for routine monitoring that they had previously struggled to adopt, despite prior clinical advice. A number of participants shared that the consistency of daily use also appeared to support habit formation over time.

Theme 4: In Addition to CHWs Assisting With Navigation of Social Barriers to Care, CHWs Provided Emotional Support and Motivation for Clinical Care Plan Adherence

Participants described CHW care as a consistent and accessible source of support during the intervention. Participants offered that CHWs assisted with helping clarify their care plans, providing accountability for managing their HF, and supporting use of the digital platform. One participant shared their CHW helped them stay on track through regular check-ins:

> *I did like having a person that I could speak with to keep me motivated - even just once a week or more if needed. I found it very helpful and supportive that there was somebody on the other end.*

A number of participants cited CHW-specific education and health care coaching as an important part of the intervention. Participants also identified CHW staff as a key resource for addressing social needs. Reported areas of support included housing navigation, transportation to appointments, insurance benefit enrollment, and access to food or home health services.

Theme 5: Connectivity Issues and Other Technical Challenges Occurred With Digital Platform Use

Some participants described specific technical challenges with digital platform use. Most issues related to device connectivity, syncing delays, or sensor functionality. For example, as one participant shared,

> *The only problem I had was on occasion the blood pressure cuff did not register a pressure easily. So it*

XSL•FO
RenderX

> *took three or four tries before a readable pressure was there, but that was the only thing.*

Other participants had intermittent issues with blood pressure measurement accuracy or precision. Participants expressed preferences to wear the arm sensor while showering for ease of use. As the digital platform's mobile app was embedded in an Android phone that needed to be carried with participants when leaving their home, some participants expressed interest in having an app that could be used on their personal phone to avoid this. Table 1 summarizes the 5 themes.

XSL•FO
RenderX

Table . Major themes and illustrative quotes from poststudy interviews with participants with heart failure (HF).

Themes	Theme description	Quotes
Theme 1: the combined intervention was well received, and CHWs[a] made the use of the digital platform more approachable.	Remote monitoring devices and app-based tools helped patients integrate health monitoring (blood pressure and weight) into daily routines, supporting adherence to care plans and fostering self-awareness of health.	• "I didn't know anything like this existed, so everything was a positive, you know what I mean? And I don't know, because I'm a little older, but sometimes I have a hard time getting information that I need. And with the app and the CHW, everything was right there. She had the answer to any question I would possibly ask about the app and it made me more willing to use it." • "Sometimes I can have a problem working with new things and new people that I don't know as well but this was easy. The app was great. She helped me get over any concerns I had about using the technology. Very comfortable and helpful overall." • "People with my condition should want to be in the study and use the devices. It will give them and their family and them a lot more security and comfort knowing that they are being tracked."
Theme 2: the intervention enhanced HF knowledge and confidence in self-care.	Participants reported gaining a better understanding of their health condition and how self-care at home was connected to their overall health.	• "I became more aware of the importance of knowing [the] baseline for my blood pressure and that my weight would fluctuate depending on what time I took my weight. So, I became more consistent with my medication." • "They helped me with a lot of stuff I really didn't know about and answered questions for me. I mean, it was nice having an advocate." • "I did find a lot of information on my blood pressure. And I'm not a person who takes my blood pressure daily, but it was good information and would frequently remind me, 'Oh, [expletive], you haven't had your medication today. What are you doing?' Because my response to the medication is kind of connected to the blood pressure numbers, so it was eye-opening for me to do that."
Theme 3: digital platform use was associated with ease of integration into daily routines.	Patients thought that the digital platform was easily integrated into their daily schedules and assisted with reminding them to adhere to clinical care plan details.	• "I've never been in the habit of weighing myself every day, even though I've had a couple of doctors say that I should weigh myself more often, but this technology really changed that." • "It makes it long-lasting in the sense that you're doing it every day. So, it's more likely going to become a habit." • "It's definitely kept me on track. I don't know that I would've gotten into the habit of weighing myself and doing my blood pressure every day if I hadn't been in this study." • "It makes it very easy to stay on track. And it makes it long-lasting in the sense that you're doing it every day. So, it's more likely going to become a habit."

XSL•FO
RenderX

Themes	Theme description	Quotes
Theme 4: in addition to CHWs assisting with navigation of social barriers to care, CHWs provided emotional support and motivation for clinical care plan adherence.	Participants felt that CHWs helped address unmet social needs, improved their understanding of how to access needed resources, and provided important social support.	• "I can be very hard person to communicate and talk with. I think I know everything. But the CHW, she educated me on the importance of the different medications and how blood pressure and my diet tie into things which was helpful. I never hung up on her or anything, so that was really good [laughter]." • "She was able to get the ball rolling with Elder Services so they can come out and do a home evaluation that I had been waiting on for a while." • "When she called, she called frequently, made sure that I understood everything that I needed to understand. And she was a very good person." • "God, I really wish I could keep the CHW forever. [laughter] I'm not going to lie. I just liked the information I was getting and I learned so much- just knowing that someone was supporting me that way was a relief." • "I thought it was great when she assisted me, especially with the Mass Health insurance application and everything." • "The CHW did everything on point. I mean, from helping with the housing and getting help with groceries. She did things that I couldn't-- I, myself, couldn't have done and I am the better for it." • "When we were talking about transportation and medical issues, she immediately, sent me an email, letting me know what programs that I qualified for which was great." • I really appreciated having help with getting the fuel assistance program. Getting access to the SNAP program also helped me. The way things are in this world today, just to be able to pay bills is important." • "With the SNAP, she got on the call with me and was able to give them the information that I didn't know I needed to have in front of me. If I have to call again, I will be ready."
Theme 5: connectivity issues and other technical challenges occurred with digital platform use.	Participants identified several improvements for the intervention focused on the digital platform.	• "I had a couple of times where I weighed myself twice or I did my blood pressure because the thing was still going around a circle after a minute or so or two, and then it didn't look like it was going to record, so I did it again." • "I found the cuff reading to vary -- it gave me a couple of very high readings that were incorrect and I discussed with CHW. Everything ended up improving after I let you guys know about it." • "It would have been great to hop in the shower and keep the sensor on, but I couldn't do that because it wasn't waterproof." • "I had to carry two phones, which it'd be nice if it could work with iOS phone instead of the android."

[a]CHW: community health worker.

XSL•FO
RenderX

Discussion

Principal Findings

This study demonstrated that digital platform use was feasible and that CHW staff assisted in making the platform more approachable for patients. This work strengthens our understanding of the barriers and facilitators to home-based HF management by identifying patient experiences with a digitally enabled CHW intervention as a part of a pilot trial. Few studies have examined patient experience with digital health interventions in the context of clinical trials. This work deepens the context of patient-centered HF home management identified in our previous studies [18,21] and other prior qualitative studies on HF [23].

Participants indicated that CHW care made the use of the digital platform more acceptable for patients. While CHW care is traditionally applied to social risk navigation [24,25], the use of CHWs specifically in digital trials for HF is novel and represents another way that CHWs can contribute to the landscape of care for patients facing HF at home [26]. Using CHW care along with HF digital platforms may help mitigate barriers to adoption and usability of digital platforms [27-29]. Despite patient hesitation regarding use of digital modalities to manage their health [30], our study findings highlighted positive participant responses regarding use of the digital platform. A recent cross-sectional study of patients with HF found that while only 22% of respondents reported prior use of health-related mobile apps, 53% of respondents expressed interest in using one for future self-management [31]. This suggests that patient interest in use of digital platforms may exceed actual use. Coimplementation of digital strategies with CHW or navigator support, even for a time-limited initiation interval for patients, could be an effective strategy for lowering barriers to digital health adoption among patients with HF.

Participants indicated that routine completion of home-based self-management skills such as monitoring home blood pressure, weight, and taking medications improved with the intervention. This is consistent with other remote monitoring studies in HF and chronic disease [32,33]. While this intervention was only 30 days in length, management skill uptake and integration into daily routines is a critical first step toward reducing preventable readmissions and improving postdischarge care transitions. One national survey of recently hospitalized patients with HF found that those who tracked their weight daily were significantly more likely to seek help for fluid retention than those who did not [34]. A number of small observational studies and trials evaluating mobile apps in HF populations have noted similar findings with improved clinician-patient communication, self-care, and engagement [35,36].

Given the established link between unmet social needs and heightened risk of HF readmission [37,38], our findings underscore the value of CHW care in efforts to help reduce risk associated with urgent and emergent care use. Participants indicated that CHW staff were helpful in assisting with navigating unmet social needs such as receiving assistance with housing, transportation, older adult services, and federal benefits navigation. By addressing health-related social needs such as those that are often associated with increased health care use, CHW care can improve care at home after discharge. These findings align with CHW core competencies and emphasize the impact of CHWs in chronic disease management [39-41]. In some cases, participants also reported learning how to engage with service agencies independently, suggesting that the intervention promoted some skill-building in this area.

We found that although participants frequently credited CHW care with enhancing their clinical knowledge of HF through personalized health care coaching, few used the digital library of educational videos associated with the mobile app. These videos were expected to be a key source of asynchronous learning and skill-based reinforcement for patients. However, despite frequent use of other aspects of the digital platform (eg, digital blood pressure cuff, weight scale, and daily questionnaire), few participants watched the videos with regularity despite encouragement from CHWs. This highlights the challenge associated with how patients access aspects of digital interventions. In retrospect, the 3-item task list for daily completion of platform activities that patients used each day did not include the videos as an itemized task. Incorporation of an option to watch educational videos as a part of the task list may augment use if adopted in future mobile app iterations.

Finally, several participants identified technical challenges related to digital platform use. Prior to the performance of this study, we conducted an open pilot trial (single arm) specifically designed to optimize workflow integration and identification of any technical issues related to the digital platform [18]. Despite the identification and reconciliation of digital platform connectivity, syncing, capture, and accuracy issues in the prior open pilot trial, participants in this study experienced a number of technical challenges. Previous studies have shown that technical glitches in remote monitoring devices can lead to decreased patient adherence and trust [42,43]. In HF-focused digital platforms, the importance of reliable operability and connectivity cannot be overemphasized [44]. Ultimately, performance accuracy and precision will decide which digital platforms merit larger-scale integration in health systems.

Limitations

This study has limitations. First, participants were recruited from a single large urban academic medical center, which may limit the generalizability of findings for patients in community or rural health systems. Second, because enrollment was limited to patients recently hospitalized for HF, the sample may reflect individuals with greater clinical and social complexity than the broader HF population. Although there are numerous studies examining digital platforms in HF cohorts [45,46] and examining CHW care in chronic disease or cardiovascular populations [47,48], the ability to discern specific effects of either the digital platform or CHW care alone within this study is limited. Although the study guide was comprehensive and thematic saturation was achieved, some patient perspectives may have still been underrepresented, particularly those of individuals who were less engaged. Finally, resource constraints limited trial enrollment eligibility to those with English proficiency, which may have excluded specific perceptions experienced by those patients fluent in languages other than

XSL•FO
RenderX

English. We also acknowledge that insights from the control group were not included here and may have provided additional findings. Despite this, we believe that these limitations were balanced by a robust study design with authentic responses gathered from participants engaged in this study.

Conclusions

This qualitative study demonstrated that patients with HF reported positive experiences with a digitally enabled CHW intervention. Most of the participants expressed that the components of the intervention worked well together and that the CHW care reduced barriers to platform use. These findings can inform digital and CHW home HF management for cardiac and primary care teams caring for this population. Further research is needed to better understand the facilitators and barriers to implementing this type of intervention in larger HF populations.

Acknowledgments

The authors would like to extend their special thanks to research study team leaders Susan Hassan, BS; Yadira Reyes-Richards, BS; and Anne Walton, RN. The authors also thank MGH Primary Care, the MGH Corrigan Minehan Heart Center, and the MGH Division of General Internal Medicine. Generative artificial intelligence tools were not used in any portion of this manuscript.

Funding

This work was supported in part by the National Institutes of Health, National Heart, Lung, and Blood Institute (1K23HL150287-01) awarded to JC. AT was also supported by the NHLBI (grant K24 HL163073). The funder had no role in the design and conduct of the study; collection management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Data Availability

Deidentified data may be obtained upon request by contacting the corresponding author with a descriptive proposal stating the purpose of the data request.

Authors' Contributions

Conceptualization: JC (lead), AT (supporting)
Methodology: JC (lead), YR-R (supporting), AT (supporting), KD (supporting)
Software: JC (lead), NS (supporting)
Validation: JC (lead)
Formal analysis: JC (lead), NS (supporting)
Investigation: JC (lead), YR-R (supporting), AT (supporting), KD (supporting)
Resources: AT (lead), KD (supporting)
Data curation: JC (lead), NS (supporting)
Writing – original draft: JC (lead), YR-R (supporting)
Writing – review and editing: JC (lead), AT (supporting), NR (supporting)
Visualization: JC (lead), NS (supporting)
Supervision: JC (lead)
Project administration: JC (lead), NS (supporting)
Funding acquisition: JC (lead)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient interview guide.
[DOCX File, 16 KB - cardio_v10i1e93288_app1.docx]

Multimedia Appendix 2
Patient characteristics.
[DOCX File, 24 KB - cardio_v10i1e93288_app2.docx]

References

1. Clark KAA, Reinhardt SW, Chouairi F, et al. Trends in heart failure hospitalizations in the US from 2008 to 2018. J Card Fail 2022 Feb;28(2):171-180. [doi: 10.1016/j.cardfail.2021.08.020] [Medline: 34534665]

XSL•FO
RenderX

2. Bozkurt B, Ahmad T, Alexander KM, et al. Heart failure epidemiology and outcomes statistics: a report of the Heart Failure Society of America. J Card Fail 2023 Oct;29(10):1412-1451. [doi: 10.1016/j.cardfail.2023.07.006] [Medline: 37797885]
3. Kazi DS, Elkind MS, Deutsch A, et al. Forecasting the economic burden of cardiovascular disease and stroke in the United States through 2050: a presidential advisory from the American Heart Association. Circulation 2024 Jul 23;150(4):e89-e101. [doi: 10.1161/CIR.0000000000001258] [Medline: 38832515]
4. Gupta A, Fonarow GC. The hospital readmissions reduction program: evidence for harm. JACC Heart Fail 2018 Jul;6(7):607-609. [doi: 10.1016/j.jchf.2018.02.012] [Medline: 29957194]
5. Osenenko KM, Kuti E, Deighton AM, Pimple P, Szabo SM. Burden of hospitalization for heart failure in the United States: a systematic literature review. J Manag Care Spec Pharm 2022 Feb;28(2):157-167. [doi: 10.18553/jmcp.2022.28.2.157] [Medline: 35098748]
6. Patel J, Rassekh N, Fonarow GC, et al. Guideline-directed medical therapy for the treatment of heart failure with reduced ejection fraction. Drugs 2023 Jun;83(9):747-759. [doi: 10.1007/s40265-023-01887-4] [Medline: 37254024]
7. Heidenreich PA, Fonarow GC, Opsha Y, et al. Economic issues in heart failure in the United States. J Card Fail 2022 Mar;28(3):453-466. [doi: 10.1016/j.cardfail.2021.12.017] [Medline: 35085762]
8. Azizi Z, Golbus JR, Spaulding EM, et al. Challenge of optimizing medical therapy in heart failure: unlocking the potential of digital health and patient engagement. J Am Heart Assoc 2024 Jan 16;13(2):e030952. [doi: 10.1161/JAHA.123.030952] [Medline: 38226520]
9. Jiang Y, Wang W. Health promotion and self-management among patients with chronic heart failure. In: Haugan G, Eriksson M, editors. Health Promotion in Health Care: Vital Theories and Research: Springer; 2021. [doi: 10.1007/978-3-030-63135-2_19]
10. Bazoukis G, Loscalzo J, Hall JL, Bollepalli SC, Singh JP, Armoundas AA. Impact of social determinants of health on cardiovascular disease. J Am Heart Assoc 2025 Mar 4;14(5):e039031. [doi: 10.1161/JAHA.124.039031] [Medline: 40035388]
11. Fonarow GC, Abraham WT, Albert NM, et al. Factors identified as precipitating hospital admissions for heart failure and clinical outcomes: findings from OPTIMIZE-HF. Arch Intern Med 2008 Apr 28;168(8):847-854. [doi: 10.1001/archinte.168.8.847] [Medline: 18443260]
12. Jacobs JA, Ayodele I, Bress AP, et al. Social determinants of health and disparities in guideline-directed medical therapy optimization for heart failure. Circ Heart Fail 2025 Jan;18(1):e012357. [doi: 10.1161/CIRCHEARTFAILURE.124.012357] [Medline: 39527624]
13. Morris A, Shah KS, Enciso JS, et al. The impact of health care disparities on patients with heart failure. J Card Fail 2022 Jul;28(7):1169-1184. [doi: 10.1016/j.cardfail.2022.04.008] [Medline: 35595161]
14. Cowie MR, Lam CS. Remote monitoring and digital health tools in CVD management. Nat Rev Cardiol 2021 Jul;18(7):457-458. [doi: 10.1038/s41569-021-00548-x] [Medline: 33824486]
15. Stevenson LW, Ross HJ, Rathman LD, Boehmer JP. Remote monitoring for heart failure management at home. J Am Coll Cardiol 2023 Jun 13;81(23):2272-2291. [doi: 10.1016/j.jacc.2023.04.010] [Medline: 37286258]
16. Koehler F, Koehler K, Deckwart O, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
17. McBeath KC, Angermann CE, Cowie MR. Digital technologies to support better outcome and experience of care in patients with heart failure. Curr Heart Fail Rep 2022 Jun;19(3):75-108. [doi: 10.1007/s11897-022-00548-z] [Medline: 35486314]
18. Hepburn J, Williams L, McCann L. Barriers to and facilitators of digital health technology adoption among older adults with chronic diseases: updated systematic review. JMIR Aging 2025 Sep 11;8:e80000. [doi: 10.2196/80000] [Medline: 40934502]
19. Carter JA, Swack N, Isselbacher E, Donelan K, Thorndike A. Feasibility, acceptability, and preliminary effectiveness of a combined digital platform and community health worker intervention for patients with heart failure: pilot randomized controlled trial. JMIR Cardio 2024 Aug 8;8:e59948. [doi: 10.2196/59948] [Medline: 38959294]
20. Carson SL, Hong C, Behforouz H, et al. Mechanisms for community health worker action on patient-, institutional-, and community-level barriers to primary care in a safety-net setting. J Ambul Care Manage 2022;45(1):22-35. [doi: 10.1097/JAC.0000000000000405] [Medline: 34812754]
21. Carter J, Donelan K, Thorndike AN. Patient perspectives on home-based care and remote monitoring in heart failure: a qualitative study. J Prim Care Community Health 2022;13:21501319221133672. [doi: 10.1177/21501319221133672] [Medline: 36305386]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Vanneste A, Barbier L, Missotten R, et al. Heart failure patients' perspectives on treatment outcomes and unmet medical needs: a qualitative preference study. ESC Heart Fail 2024 Oct;11(5):3075-3084. [doi: 10.1002/ehf2.14891] [Medline: 38856019]

XSL•FO
RenderX

24. Larson EK, Ingram M, Dougherty E, et al. Centering the role of community health workers in social risk screening, referral, and follow-up within the primary care setting. BMC Prim Care 2024 Sep 13;25(1):338. [doi: 10.1186/s12875-024-02590-3] [Medline: 39271996]
25. Carter J, Hassan S, Walton A. Meeting the needs of vulnerable primary care patients without COVID-19 infections during the pandemic: observations from a community health worker lens. J Prim Care Community Health 2022;13:21501319211067669. [doi: 10.1177/21501319211067669] [Medline: 35068240]
26. Feroz AS, Khoja A, Saleem S. Equipping community health workers with digital tools for pandemic response in LMICs. Arch Public Health 2021 Jan 4;79(1):1. [doi: 10.1186/s13690-020-00513-z] [Medline: 33390163]
27. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
28. Frishammar J, Essén A, Bergström F, Ekman T. Digital health platforms for the elderly? Key adoption and usage barriers and ways to address them. Technol Forecast Soc Change 2023 Apr;189:122319. [doi: 10.1016/j.techfore.2023.122319]
29. Zaman SB, Khan RK, Evans RG, Thrift AG, Maddison R, Islam SM. Exploring barriers to and enablers of the adoption of information and communication technology for the care of older adults with chronic diseases: scoping review. JMIR Aging 2022 Jan 7;5(1):e25251. [doi: 10.2196/25251] [Medline: 34994695]
30. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
31. Leigh JW, Gerber BS, Gans CP, Kansal MM, Kitsiou S. Smartphone ownership and interest in mobile health technologies for self-care among patients with chronic heart failure: cross-sectional survey study. JMIR Cardio 2022 Jan 14;6(1):e31982. [doi: 10.2196/31982] [Medline: 35029533]
32. Lee SG, Blood AJ, Cannon CP, et al. Remote cardiovascular hypertension program enhanced blood pressure control during the COVID-19 pandemic. J Am Heart Assoc 2023 Mar 21;12(6):e027296. [doi: 10.1161/JAHA.122.027296] [Medline: 36915035]
33. Blood AJ, Cannon CP, Gordon WJ, et al. Results of a remotely delivered hypertension and lipid program in more than 10 000 patients across a diverse health care network. JAMA Cardiol 2023 Jan 1;8(1):12-21. [doi: 10.1001/jamacardio.2022.4018] [Medline: 36350612]
34. Tong CW, Kanwal A, Sharma GV, Bither C, Walsh MN, Byrd K, et al. Acute decompensated heart failure: the need for the patient's perspective. Cardiology Magazine. 2020 Jul 23. URL: https://www.acc.org/latest-in-cardiology/articles/2020/07/01/12/42/acute-decompensated-heart-failure-the-need-for-the-patients-perspective [accessed 2026-05-27]
35. Bezerra Giordan L, Ronto R, Chau J, Chow C, Laranjo L. Use of mobile apps in heart failure self-management: qualitative study exploring the patient and primary care clinician perspective. JMIR Cardio 2022 Apr 20;6(1):e33992. [doi: 10.2196/33992] [Medline: 35442205]
36. Azizi Z, Broadwin C, Islam S, et al. Digital health interventions for heart failure management in underserved rural areas of the United States: a systematic review of randomized trials. J Am Heart Assoc 2024 Jan 16;13(2):e030956. [doi: 10.1161/JAHA.123.030956] [Medline: 38226517]
37. Sterling MR, Ringel JB, Pinheiro LC, et al. Social determinants of health and 30-day readmissions among adults hospitalized for heart failure in the REGARDS study. Circ Heart Fail 2022 Jan;15(1):e008409. [doi: 10.1161/CIRCHEARTFAILURE.121.008409] [Medline: 34865525]
38. Enard KR, Coleman AM, Yakubu RA, Butcher BC, Tao D, Hauptman PJ. Influence of social determinants of health on heart failure outcomes: a systematic review. J Am Heart Assoc 2023 Feb 7;12(3):e026590. [doi: 10.1161/JAHA.122.026590] [Medline: 36695317]
39. Covert H, Sherman M, Miner K, Lichtveld M. Core competencies and a workforce framework for community health workers: a model for advancing the profession. Am J Public Health 2019 Feb;109(2):320-327. [doi: 10.2105/AJPH.2018.304737] [Medline: 30571307]
40. Kangovi S, Mitra N, Grande D, Long JA, Asch DA. Evidence-based community health worker program addresses unmet social needs and generates positive return on investment. Health Aff (Millwood) 2020 Feb;39(2):207-213. [doi: 10.1377/hlthaff.2019.00981] [Medline: 32011942]
41. Jack HE, Arabadjis SD, Sun L, Sullivan EE, Phillips RS. Impact of community health workers on use of healthcare services in the United States: a systematic review. J Gen Intern Med 2017 Mar;32(3):325-344. [doi: 10.1007/s11606-016-3922-9] [Medline: 27921257]
42. Taylor ML, Thomas EE, Vitangcol K, et al. Digital health experiences reported in chronic disease management: an umbrella review of qualitative studies. J Telemed Telecare 2022 Dec;28(10):705-717. [doi: 10.1177/1357633X221119620] [Medline: 36346938]
43. Main A, McCartney H, Ibrar M, et al. Patients' experiences of digital health interventions for the self-management of chronic pain: systematic review and thematic synthesis. J Med Internet Res 2025 Mar 18;27:e69100. [doi: 10.2196/69100] [Medline: 40101209]
44. Castiello T, Hill L, Man S, Elseidy S, Griffin D, Schuuring MJ. Digital solutions in HF education: what can patients and clinicians gain? Curr Heart Fail Rep 2026 Jan 9;23(1):4. [doi: 10.1007/s11897-025-00738-5] [Medline: 41511585]

XSL•FO
RenderX

45. Brahmbhatt DH, Ross HJ, O'Sullivan M, et al. The effect of using a remote patient management platform in optimizing guideline-directed medical therapy in heart failure patients: a randomized controlled trial. JACC Heart Fail 2024 Apr;12(4):678-690. [doi: 10.1016/j.jchf.2024.02.008] [Medline: 38569821]
46. Johnson AE, Routh S, Taylor CN, et al. Developing and implementing an mHealth heart failure self-care program to reduce readmissions: randomized controlled trial. JMIR Cardio 2022 Mar 21;6(1):e33286. [doi: 10.2196/33286] [Medline: 35311679]
47. Krantz MJ, Coronel SM, Whitley EM, Dale R, Yost J, Estacio RO. Effectiveness of a community health worker cardiovascular risk reduction program in public health and health care settings. Am J Public Health 2013 Jan;103(1):e19-e27. [doi: 10.2105/AJPH.2012.301068] [Medline: 23153152]
48. Allen JK, Himmelfarb CR, Szanton SL, Bone L, Hill MN, Levine DM. COACH trial: a randomized controlled trial of nurse practitioner/community health worker cardiovascular disease risk reduction in urban community health centers: rationale and design. Contemp Clin Trials 2011 May;32(3):403-411. [doi: 10.1016/j.cct.2011.01.001] [Medline: 21241828]

Abbreviations

CHW: community health worker
HF: heart failure
REDCap: Research Electronic Data Capture

Edited by A Coristine; submitted 15.Mar.2026; peer-reviewed by S Turnbull; revised version received 16.May.2026; accepted 18.May.2026; published 04.Jun.2026.

Please cite as:
Carter J, Swack N, Rehman N, Reyes-Richards Y, Donelan K, Thorndike A
Patient Perceptions of a Digitally Enabled Community Health Worker Intervention: Qualitative Study Among Pilot Trial Participants
JMIR Cardio 2026;10:e93288
URL: https://cardio.jmir.org/2026/1/e93288
doi:10.2196/93288

© Jocelyn Carter, Natalia Swack, Narmeen Rehman, Yadira Reyes-Richards, Karen Donelan, Anne Thorndike. Originally published in JMIR Cardio (https://cardio.jmir.org), 4.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The HeartHealth Program: A Mixed Methods Study of a Community-Based Text Messaging Support Program for Patients With Cardiovascular Disease From 2020 to 2024

Brodie Sheahen[1], MD; Liliana Laranjo[1], MD, PhD; Ritu Trivedi[2], PhD; Tim Shaw[1,3], PhD; Gopal Sivagangabalan[1,4], MBBS, PhD; James Chong[1,4], MBBS, PhD; Aravinda Thiagalingam[1,4], MBBS, PhD; Sarah Zaman[1,4], MBBS, PhD; Pierre Qian[1,4], MBBS, PhD; Anupama Balasuriya Indrawansa[1], MBBS; Clara Kayei Chow[1], MBBS, PhD

[1]Westmead Applied Research Centre, The University of Sydney, Level 6, Block K, Entrance 10, Westmead Hospital, Hawkesbury Rd, Sydney, Australia

[2]School of Health Sciences, The University of Sydney, Sydney, Australia

[3]Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia

[4]Department of Cardiology, Westmead Hospital, Sydney, Australia

Corresponding Author:
Brodie Sheahen, MD
Westmead Applied Research Centre, The University of Sydney, Level 6, Block K, Entrance 10, Westmead Hospital, Hawkesbury Rd, Sydney, Australia

Abstract

Background: The HeartHealth program is a 6-month SMS text messaging–based support program offered to patients with a recent cardiovascular hospitalization or recent cardiovascular clinic visit in Western Sydney, Australia. Its customized content focuses on cardiovascular risk factors, lifestyle, treatments, and general heart health information.

Objective: This study aimed to evaluate the implementation of the HeartHealth program.

Methods: A mixed methods study was conducted assessing program reach, effectiveness, implementation, and maintenance using program data, participant feedback surveys, and staff focus group discussions. Consecutive adult patients who had attended cardiology clinics or had been discharged from cardiology hospitalization at Westmead Hospital, between April 2020 and April 2024, were included in the analysis. Content analysis was used to interpret the qualitative data.

Results: A total of 23,095 patients were invited, 8804 (38.1%) enrolled into the program, and 7964 out of 8804 (90.5%) completed the 6-month duration. Participants enrolled in the HeartHealth program had a mean age of 58.6 years, 60.3% (5302/8788) were male, and 62.4% (5382/8624) were recruited from an outpatient clinic setting. A total of 851,058 SMS text messages were sent, with 99.41% (846,009/851,058) delivered successfully. A total of 3533 out of 7964 (44.4% of program completers) participants completed the postintervention survey, and 4 HeartHealth staff members participated in a focus group discussion. Among the participants who completed the survey, 60.5% (2137/3533) reported that the program improved the healthiness of their diet, 53.6% (1894/3533) reported improved physical activity levels, and 56.1% (1982/3533) reported that it helped remind them to take their medications. Content analysis of participant feedback identified that the program was effective in prompting participants to change their diet, providing emotional support, reminding them of the importance of behavior change, improving their confidence in managing their health, and keeping participants focused. Key barriers included limited personalization, language options, and SMS text messaging scheduling flexibility. Recommended adaptations focused on enhancing personalization, greater engagement by local clinical teams, and expanding program dissemination.

Conclusions: The program had a broad reach, translated to improved patient-reported health behaviors, and provided participants with needed support at low cost and low resource requirements. This analysis highlights the successful implementation and scalability of the HeartHealth program and provides key learnings for health systems that are looking to implement similar programs in the future.

(JMIR Cardio 2026;10:e68896) doi:10.2196/68896

KEYWORDS

text messaging; Short Message Service; SMS; secondary prevention; implementation evaluation; cardiovascular disease; digital health intervention

XSL•FO
RenderX

Introduction

Patient education is an important element of care of patients with chronic disease, and several international agencies now have a policy focus on improving patient education to support self-management [1-3]. Smoking cessation, physical activity, diet modification, weight management, and medication management are all important in cardiovascular disease (CVD) prevention, but current models of care are limited in support for patients to address these behaviors [4]. Most current health care models' education and support are typically delivered through in-person consultation and group activities; however, this modality poses multiple barriers such as many travel requirements [5], cost [6], time, and lack of prioritization [7]. In addition, many health care services lack the resources and funding to deliver these, and programs have limited ability to cater to population diversity [8,9].

Digital health technologies present a scalable means to deliver customized patient education [10]. Several small- to medium-sized randomized controlled trials have shown mobile health texting interventions to improve patient CVD risk factors, including: low-density lipoprotein cholesterol levels [11], BMI [11], blood pressure [12], weight management [13], and smoking cessation [14]. Furthermore, these interventions report high rates of patient satisfaction [15] and usability [11]. Yet, despite the research to date, there are limited examples of real-world implementation and evaluation of large-scale digital education and support programs for cardiovascular and other chronic diseases.

In April 2020, the HeartHealth program was initiated and offered to patients with CVD discharged from cardiovascular services or clinics in Western Sydney, New South Wales. The program provides personalized cardiovascular education and support via SMS text messaging through a digital customization platform over a 6-month period and an opportunity to ask questions. This study aimed to evaluate the implementation of the HeartHealth program.

Methods

HeartHealth Program Description

The HeartHealth program was designed to reduce cardiovascular risk in patients with CVD or at high CVD risk. The HeartHealth program involved the delivery of regular semipersonalized cardiovascular education and support via SMS text message for 6 months. SMS text messages were sent approximately 3 - 4 times per week. The message bank was developed by clinicians, academics, and patients and covered the following 5 modules: smoking, diet, physical activity, COVID-19, and general cardiovascular health. Messages were written to provide advice, education, motivation, and reminders aimed at improving cardiovascular risk factors and healthy lifestyle behaviors. SMS text messages would often be supplemented with a URL link to a website to enable access to further information on the message content. Participants were able to opt out of the program at any time through alerting staff through responding to the SMS text messages. The core structure of the program content was curated based on the previously published TextMe and TextMe2 programs [11,16]. Message content development was based on a range of theoretical frameworks spanning 3 phases of development as previously described [17,18], with program content actively reviewed and updated.

At registration, participants completed a survey detailing their baseline characteristics (hypertension, diabetes mellitus, hypercholesterolemia, smoking status, and diet preference). Algorithms selected messages from the message bank based on participants' baseline characteristics, tailoring each program accordingly. Messages addressed the participants by their preferred name and provided the source of information; examples of SMS text messages have previously been outlined in program development protocols [17,18]; furthermore, Multimedia Appendix 1 provides examples of messages.

To disseminate SMS text messaging support programs at scale, our team at Westmead Applied Research Centre built a cloud-based digital platform "TextCARE" that can deliver multiple programs according to varying clinical algorithms, simultaneously. Hence, this enabled delivery of customized content to thousands of people concurrently. The HeartHealth program that started at Westmead Hospital in April 2020 continues to be deployed.

Enrollment and Eligibility

Patients were identified either following attendance at a Westmead Hospital outpatient rapid access cardiology clinic or following discharge from an inpatient cardiology admission at Westmead Hospital. Patients were eligible for HeartHealth if they were aged 18 years or older, post hospital discharge from a cardiology admission, or had recently attended an outpatient cardiology clinic. Initially, the program was designed where consecutive eligible patients from the previous week were sent a single SMS text message to enroll into the HeartHealth program. Recruitment protocols were adapted in October 2021 so that participants who did not respond to the initial SMS text messaging invitation were followed up with a phone call from a HeartHealth staff member and offered program enrollment. Consent was obtained electronically, disseminated by an SMS text message, and captured on REDCap (Research Electronic Data Capture; Vanderbilt University).

Ethical Considerations

This study was approved by the Western Sydney Local Health District Human Research Ethics Committee (approval number 2020/ETH01649). Consent for data collection was provided by participants at the time of program enrollment. All data collected are deidentified. There was no compensation provided to participants for their involvement with this study.

Study Design

This study is a retrospective observational study that evaluated the implementation of the HeartHealth program, an existing program implemented as standard of care in the Western Sydney Local Health District, New South Wales, Australia. A mixed methods design assessing the HeartHealth program "Reach," "Effectiveness," "Implementation," and "Maintenance" was used. Typically, an implementation evaluation would also assess program site "adoption"; however, as the HeartHealth program

XSL•FO
RenderX

was intended to be rolled out only at one site, this component was not assessed. Three sources of data were collected: postintervention surveys, focus group discussions with organization staff, and program-related data. Data sources used for analysis are outlined in Table 1.

Table . Descriptions of the implementation evaluation domains including the domain definition, outcome measures, and data sources.

Domain	Definition	Outcome measures	Data sources
Reach	Reach was defined as the characteristics and proportion of patients who agreed to opt into the HeartHealth program following hospital discharge or outpatient clinic visit		
		• Participant enrollment	• Program data[a]
		• Characteristics of patients who did and did not opt into the program	• Program data • Electronic medical health records (adapted with permission from Sheahen et al., [19])
		• Reasons patients did not opt into the program	• Program data
Effectiveness	Effectiveness was defined as clinical improvements following the program delivery and was assessed by participant behavior and knowledge changes		
		• Participant lifestyle behavior changes	• Postintervention survey
		• Participant health knowledge and behavior.	• Postintervention survey
Implementation and maintenance	Implementation and maintenance were defined as the extent the staff members implemented and maintained the program as intended as well as the participant's perception of appropriate program delivery		
		• Program fidelity, attrition, and organization requirements	• Program data
		• Program resources and costs	• Program data
		• Participant SMS text message interaction	• Program data
		• Program barriers (organization and individual levels)	• Focus group discussion • Postintervention survey
		• Program enablers (organization and individual levels)	• Focus group discussion • Postintervention survey
		• Program adaptations required for long-term maintenance (organization and individual levels)	• Postintervention survey • Focus group discussion
		• Program adaptations made by HeartHealth staff	• Focus group discussion • Program data

[a]Data obtained from participants and staff members during enrollment, throughout the program, and at program completion as per the programs standard practice.

XSL•FO
RenderX

Data Collection

Postintervention Surveys

All participants in the HeartHealth program were invited to complete an assessment survey at the end of the intervention (Multimedia Appendix 2). The survey was designed by The University of Sydney staff to evaluate the program, asking participants for feedback on the program's impact, such as changes in lifestyle behaviors, what they enjoyed, and what could be improved. The survey was distributed via REDCap.

Focus Group Discussion

All current HeartHealth staff members (n=4) involved with the operationalization of the HeartHealth program were invited to partake in a single focus group discussion regarding the reach, effectiveness, implementation, and maintenance of the HeartHealth program (Multimedia Appendix 3).

HeartHealth Program Data

From the onset of the HeartHealth program, staff members recorded and stored data on participant outreach, enrollment, opt-out, and responses to SMS text messages. On a weekly basis, staff members would record, categorize, and store the program data, enabling thorough and complete analysis on the reach and implementation components of our analysis. These data were stored securely on The University of Sydney Research Data Store platform.

Data or Statistical Analysis

Statistical analysis was undertaken using R statistical software (version 4.2.0; R Core Team). Categorical data, including quantitative survey data, program attrition data, and participant demographic data, are presented as frequencies and percentages. Qualitative data assessing participant and staff perspectives of program barriers, enablers, areas of required adaptations, interaction with SMS text messages, and implementation of the program are analyzed via content analysis [20]. One researcher (BS) familiarized themself with the data and inductively coded the data into themes and subthemes. Three researchers (BS, RT, and LL) then reviewed and discussed established themes and subthemes, which were repeatedly adapted until agreement was reached on final theme and subtheme formation. HeartHealth participants were excluded from the "Effectiveness" section of the analysis if they did not complete the entire 6-month program duration.

Results

Reach and Participant Enrollment

A total of 23,095 patients, who had either attended a Westmead Hospital cardiology clinic or were discharged from Westmead Hospital cardiology unit between April 2020 and April 2024, were offered the HeartHealth program. A total of 14,291 patients did not opt into the program, and 8804 patients consented to participate in the program (enrollment rate: 8804/23,095, 38.1%; Figure 1).

Most patients were invited to enroll in the program following an outpatient clinic review (59.7% vs 40.3%); consequently, the HeartHealth program cohort primarily consisted of patients recruited from a clinic setting. HeartHealth participants were slightly younger (58.6 years vs 61.7 years) than the nonenrollee patients; however, gender distribution was similar (Table 2). Extracted electronic medical record data on comorbidities were obtained for 4324 HeartHealth participants and 7218 nonenrollee patients from April 2020 to April 2022. During this period, program participants were more likely to report English as their preferred language than nonenrollee patients (82.1% vs 76.0%, respectively) and less likely to have prior CVDs than nonenrollee patients (eg, ischemic heart disease 15.2% vs 20% and heart failure 25.4% vs 32.7%, respectively; Multimedia Appendix 4—adapted with permission from Sheahen et al [19]).

XSL•FO
RenderX

Figure 1. HeartHealth program enrollment schema outlining the number of patients invited, enrolled, and completed the postintervention survey.

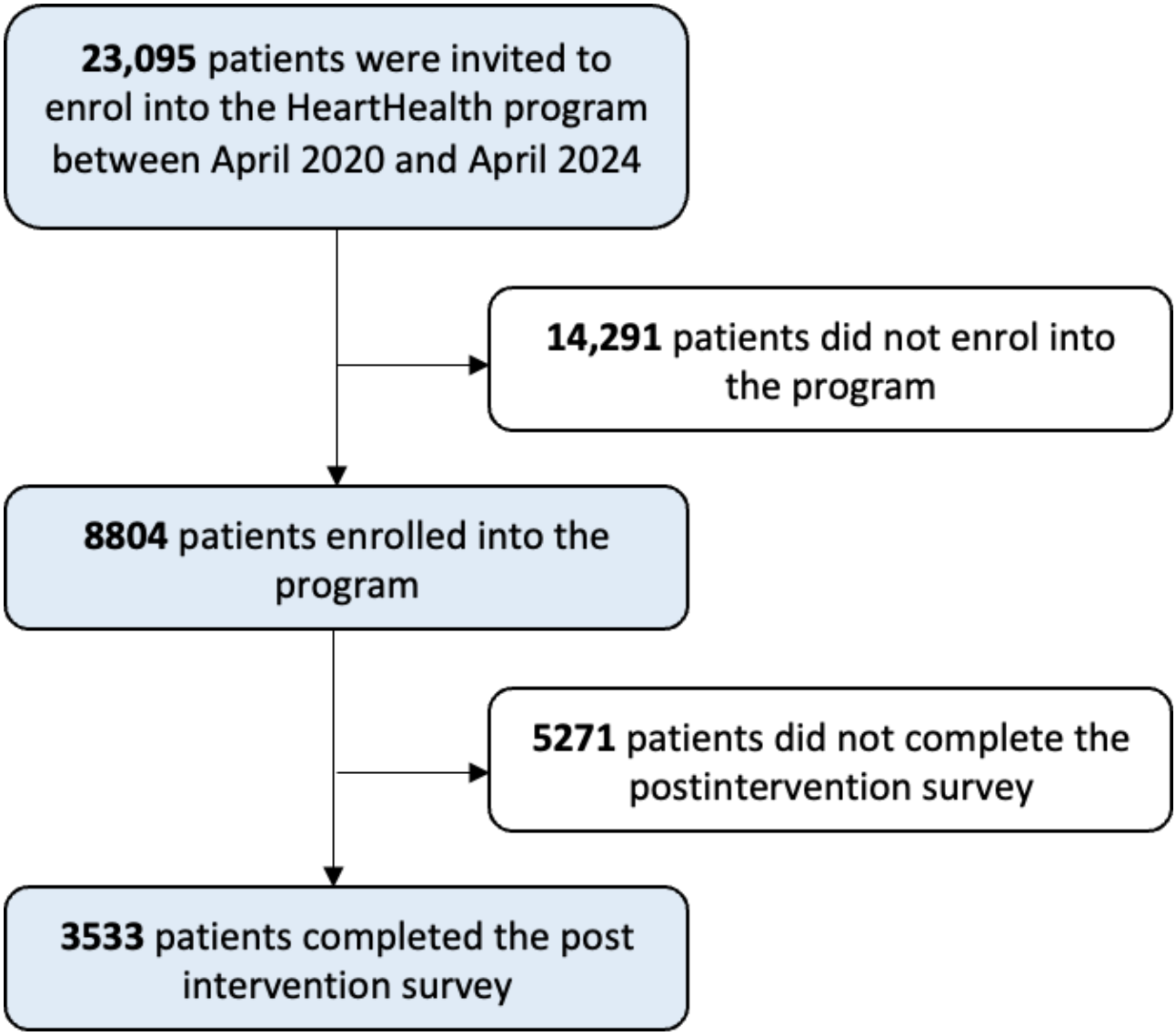


Table . Characteristics of participants who opted into the HeartHealth program and those who did not opt in (nonenrollee patients), including age, gender, and site of recruitment.

Characteristics	HeartHealth participants (n=8804)	Nonenrollee patients (n=14,291)	Overall (N=23,095)
Age (years), mean (SD)	58.6 (16.1)	61.7 (17.8)	60.5 (17.2)
Missing, n	12	10	22
Sex			
Male, n/N (%)	5302/8788 (60.3)	8647/14,263 (60.6)	13,949/23,051 (60.5)
Female, n/N (%)	3486/8788 (39.7)	5616/14,263 (39.4)	9102/23,501 (39.5)
Missing, n	16	28	44
Site of recruitment			
Hospital setting, n/N (%)	3242/8624 (37.6)	5872/14,013 (41.9)	9114/22,637 (40.3)
Clinic setting, n/N (%)	5382/8624 (62.4)	8141/14,013 (58.1)	13,523/22,637 (59.7)
Missing, n	180	278	458

Effectiveness

A total of 3533 out of 7964 (44.4% of participants who completed the program) participants completed the postintervention survey. Of those participants who completed the survey, most participants reported that the program helped them improve the healthiness of their diet (n=2137, 60.5%), increase their exercise levels (n=1894, 53.6%), and remind them

XSL•FO
RenderX

to take their medications (n=1982, 56.1%) [1]. Content analysis of participant feedback identified that HeartHealth participants reported that the program was effective in improving their overall health (Textbox 1). The first theme highlighted that the program translated into improved health behaviors through initiating lifestyle changes, empowering patients to engage in self-management of their health, and reinforcing healthy behaviors. The second theme demonstrated that the program improved patients' well-being and psychological health through improving patients' sense of care, accountability, motivation, and positivity.

XSL•FO
RenderX

Textbox 1. Themes and subthemes on participant benefits of the HeartHealth program.

Theme 1: Program translated to effective health behavior change

Initiating lifestyle change

- "Extremely useful [HeartHealth program], I've made changes to my lifestyle and feel I'm taking more control over myself."
- "Thanks to your support I am now on the right path to getting better, I first thought I wouldn't make next year . By slowly following the beginning of program and taking it slowly day by day . I didn't have any blockages in my heart, but my [myocardial infarction] was caused through either trauma or stress, so I naturally sourced methods from your program to help me."
- "The program has prompted me to reduce my salt intake. While I already knew that salt should be reduced in our diets, it was useful being prompted about it and having suggestions on how to reduce salt intake."
- "I was overweight, I didn't eat [many] vegetables, but now I do. I didn't walk much, now I walk every day . This program has changed my [lifestyle]."
- "I was pleased to be part of the trial which was a big factor in me losing 20 KG and becoming fitter."
- "The only unhealthy thing about my lifestyle is smoking cigarettes, those text reminders did make me think about quitting though – so I guess that's a positive."
- "The smoking messages encouraged me to quit smoking. 3 months [since I] quit."

Empowering people to engage in self-management

- "I found all the messages were very useful for me as it empowered me to do more activities, eat healthy, take my tablets and to stay healthy. I'm well and improved a lot beyond expectation. Thank you so much."
- "After following this program I am confident I can follow through with the activities suggested and have healthy foods in the future for improvement to my physical health and internal health."
- "Most of the messages are useful and doable. I have gained a lot of important information necessary to maintain a healthy lifestyle."

Reinforcement of healthy habits

- "Short and sharp reminders helped me re-focus especially with regard to diet and exercise."
- "Good work. I'm glad I'm doing this program. Keeps me more conscious of my lifestyle and what needs improvement for a longer and healthier life."
- "The messages keep me focused, and I review them weekly just to make sure I am on track."

Theme 2: Program supported patients' psychological health and well-being

Provided participants with a sense of care

- "I love the program. I'm single and live alone and it was nice to receive a message with helpful advice and it made me much less stressed, and I felt less alone."
- "Was nice to get messages to know someone was caring and supportive through my new transitional time."
- "The messages made me feel like there was someone checking on me and was steering me in the right direction."
- "Great program. I felt that I had daily support/companion to manage my condition."

Participant accountability

- "Keep it going. It helped me so much and made me feel that I was cared for, there was someone looking out for me and it kept me on my toes."
- "It was a good program. It kept me informed of that I was supposed to do to stay healthy. A weekly reminder, at least, does not let you forget your obligations."

Motivation and positivity

- "The messages were good motivation. It's easy to revert away from a healthier lifestyle and forget about my condition but the but the messages kept it front of mind."
- "I was glad to do it. It motivated me a lot. Kept on straight and narrow. I looked forward to receiving the massages. Thank you."
- "The increase in psychological safety was a direct result. This helped in keeping me focused and in a positive state of mind as I journeyed through regaining my health and amending my lifestyle."

XSL•FO
RenderX

Implementation and Maintenance: Fidelity and Attrition

A total of 851,058 (average 97 per participant) SMS text messages were sent between April 21, 2020, and April 1, 2024; 99.41% (846,009/851,058) were successfully delivered. Of the 8804 patients who participated in the HeartHealth program, 9.5% (n=840) did not complete the 6-month program, 25.0% (210/840) withdrew from the program in the first week, 52.9% (444/840) withdrew between weeks 2 and 13, and 22.1% (186/840) withdrew between weeks 14 and 26. In comparison with those patients who completed the intervention, the group of participants who did not complete the 6-month intervention was younger (mean 56.89, SD 18.7 years vs mean 58.8, SD 15.8 years; *P*=.003), more likely to be female (42.5% vs 39.2%; *P*=.09), and had a lower prevalence of cardiovascular risk factors (hypertension 38.3% vs 48.9%, *P*<.001; hypercholesterolemia 35.2% vs 44%, *P*=.002; and diabetes mellitus 17.9% vs 24.3%, *P*<.001).

Resources and Costs

The program was operationalized by 4 The University of Sydney—Westmead Applied Research Centre staff members on a part-time basis; these roles included program manager, digital product manager, health administrator, and research assistant, each with respective responsibilities as outlined in Multimedia Appendix 5. The resources and corresponding expenses required to operationalize the HeartHealth program over the 4 years were $276,728.36 (at the time of study analysis period completion [April 1, 2024], Aus $ to US $ was $0.65. Based on this, conversion to US $ is $179,873.43 [$20.43 per participant]). Since we initially provided this as an SMS text messaging–only recruitment strategy, and then later provided it as an SMS text messaging and phone call recruitment strategy, we have provided the separate estimated costs for each of these recruitment strategies in Multimedia Appendix 6. The estimated costs using the SMS text messaging–only recruitment strategy were $75,809.13 (at the time of study analysis period completion [April 1, 2024], Aus $ to US $ was $0.65. Based on this, conversion to US $ is $49,275.93 [$19.93 per participant]), and the estimated costs using the SMS text messaging and phone call recruitment strategy were $200,919.23 (at the time of study analysis period completion [April 1, 2024], Aus $ to US $ was $0.65. The cost converted to US $130,597.50 [$20.63 per participant]). A total of 2473 participants were recruited over the 18 months of the SMS text messaging–only recruitment period (137 per month), whereas a total of 6331 participants were recruited over the 30 months of the SMS text messaging and phone call period (211 per month).

Participant SMS Text Message Interaction

Participants could reply to the messages or ask questions; it was not actively encouraged to reply, but program staff did monitor SMS text messages and respond as necessary. Across a total of 8804 people enrolled between April 2020 and April 2024, a total of 8288 responses were received. In total, 73% (6050/8804) of these responses were expressing thanks or acknowledging receipt of the message sent to them. The other 27% (2238/8288) of responses were based on lifestyle behaviors or administrative content.

Barriers and Enablers to Implementing the HeartHealth Program

From a participant perspective, there were 2 main themes elicited as program enablers. First, the program content was valuable and appropriate, as it reinforced existing knowledge, improved cardiac health awareness, and improved engagement by providing relatable and actionable information. Second, the program was communicated in an effective manner, where the frequency of the messages provided steady reminders to participants and the URL hyperlinks allowed access to further information (Textbox 2). In contrast, the barriers identified by some participants were that information was not personalized enough, there were limited language options offered to participants, there was a lack of flexibility in the message delivery timing, and the content, at times, was overly simplistic (Multimedia Appendix 7).

Adaptations Made When Implementing the HeartHealth Program

From a HeartHealth staff perspective, program feedback aligned with participant survey feedback on the perceived program use, the benefit of SMS text messaging personalization, and ease of program use. In addition, staff felt empowered and effective in their ability to manage and communicate with participants if issues arose (Multimedia Appendix 8). The main barriers perceived were lack of promotion of the program by local clinical staff, leading to participants being unaware of the program and the limited digital health literacy or English literacy of some participants, thereby requiring assistance from staff or family to onboard them to the program (Multimedia Appendix 9). Recommended future program adjustments were focused around overcoming these barriers, expanding program dissemination, and improving program personalization (Multimedia Appendix 10).

Three key adaptations were made by the HeartHealth staff following program commencement. First, efforts were made to increase “site staff program awareness” through emailing senior medical staff, presenting at departmental meetings, providing additional verbal and written education to all staff, and placing posters in wards, clinics, and frequented locations. Second, HeartHealth staff members made adaptations to “improve the enrollment process”; these adaptations aimed to improve patient understanding of the program and ease of enrollment through simplifying the initial enrollment SMS text message and implementing a follow-up phone call to assist in this process where required. In total, 19.1% (1390/7281) of the follow-up phone calls resulted in patients enrolling in the program during the call and 15.3% (1114/7281) enrolling following the call. Third, the SMS text message content was continually adapted throughout the program to provide up-to-date information on COVID-19, guidelines, and cardiovascular health (Multimedia Appendix 11).

XSL•FO
RenderX

Textbox 2. Participant-reported enablers for implementing the HeartHealth program.

Theme 1: Appropriate and valuable content was delivered

SMS text messaging content reinforced existing knowledge

- "Most of the information I knew but liked getting the messages to reinforce my knowledge."
- "All the texts were very helpful and from memory, some were repeated which was fantastic as it kept reinforcing the message if I had not taken on board what the message was telling me."

SMS text messaging content provided awareness of their cardiac condition

- "Messages were a reminder that despite feeling well I still have a chronic heart disease and need to take care about that."
- **"**The message content wasn''t the most important to me as was the reminder that I did need to consider my cardio health overall."
- "The messages provided additional reminder/reinforcement of the need to pay regular attention to aspects of lifestyle that affect health and wellbeing particularly in the context of my medical conditions."
- "For a brief moment every day I was reminded to do all I could for my health. My heart is of great concern to me and I have learnt a bit with the resources you have recommended."

Engaged with relatable information

- "My favourite messages that explain a little about the science of heart conditions and how the text advice could assist with that."
- "My favourite message was the one that described what happens to the body as a result of exercise in video format."
- "Symptoms of a heart attack or related heart disease the viewer could have & what to do.
- "More points on how smoking alcohol and eating the wrong foods can damage your heart and body."

Engaged with actionable advice

- "I liked the ideas, suggestions and reminders to get me thinking about what I could do more than the more-prescriptive messages."
- "Diet messages about nuts and salad to avoid sugar and reduce cholesterol are very encouraging."
- "They were all good, some better than others. I liked the ones with actionable [information]."

Theme 2: Information was communicated in an effective manner

Frequency of SMS text messages serving as consistent reminders

- "I have been to Weight Watchers many times and know what I should do to keep healthy but unless you are getting somebody or a text message every day , you go backwards. The daily reinforcement is the key to my success. I want to thank you all for allowing me to be a participant as you have certainly made a difference in keeping me healthy and happy."
- "Extremely important educational information...However, the frequent texts reminders help to be mindful on following the diet, medication and exercise plan. Many thanks."
- "The actual message was less important than the fact that they reminded me to be careful of diet and to exercise regularly."

Hyperlinks facilitated expanding knowledge

- "My favourite messages were 'healthy type of facts' or ideas with hyperlinks. [This allowed] the recipient the option to investigate further."
- "I was interested in the messages that provided links to more detailed and comprehensive advice, particularly about salt intake."
- "In general the comments where helpful, however, some of the messages might have links for further help. E.g., There was a message which suggested using herbs to add flavour to reduce salt. Finding the relevant information was very hard to find. Actual suggestions or a link would have been a lot more helpful."

Discussion

Principal Results

This paper describes the initial implementation and a detailed appraisal of an algorithm-driven personalized digital education and support program "HeartHealth" for patients with heart conditions. Key learnings of this study were that (1) the program was able to be implemented with high fidelity with relatively low-resource usage; (2) the majority of participants completed the 6-month program; however, program noncompletion was more commonly seen in patients of a younger age, female sex, and a lower prevalence of CVD risk factors; (3) most participants who completed the postintervention survey reported improved health and behavioral risk factors; (4) content analysis of feedback questionnaires indicated that program benefits were driven by improved self-efficacy, feeling psychologically supported, and initiating healthy lifestyle behaviors; and (5) further personalization and further engagement with local stakeholders could improve engagement and impact of the program.

XSL•FO
RenderX

The program had a high reach of the target population who were offered enrollment within a 4-year time frame. The high rates of enrollment were likely facilitated by the simple method of enrollment and that patients may have been motivated to address their heart health because of their recent hospitalization or clinic visit. It was notable that participants, compared with nonenrollee patients, were younger and, consistent with this, had fewer comorbidities. Recent studies that followed a similar program structure by automatically obtaining eligible patient contact details were also able to reach a large number of their targeted patients [21,22]. Opt-out models in cardiovascular rehabilitation programs have also shown significant increases in patients referred to the program compared with opt-in models [23,24]. Future adaptations to the HeartHealth program are required to optimize enrollment rates; using an opt-out model may be an effective option. Telephone detailing could increase enrollment and may be particularly helpful for older participants with more comorbidities. Our previous analyses have indicated that the HeartHealth program may be more effective in reducing hospitalization in older participants and could justify the additional resources for telephone detailing [19].

The content analysis of participants provides multiple insights into the reasons for improved health effects from the program. These factors included a combination of direct and indirect factors, such as the psychological or emotional support, continued light-touch connectivity, and increased self-efficacy that improved patient experiences and also encouraged positive behavior changes. These reasons align with prominent psychobehavioral theories such as the theory of planned behavior [25], social cognitive theory [26], and the self-determination theory [27], which should be further used when planning future program adaptations and dissemination. It is well recognized that providing emotional support is essential for patient-centered care [28,29] and when combined effectively with clinical care, it positively increases the patient experience [30-32]. Through the HeartHealth program providing an avenue of continual support via the SMS text messaging platform, participants were highly engaged with the program. Importantly, many replies were a general comment or to say “thanks,” rather than using the program as a modality to report health concerns requiring a reply from the health counselor, which is consistent with previous research [33]. This form of engagement provides patients with an increased sense of care and support without being resource-intensive and demanding on staff workload. To continue delivering an effective program, it is vital to consider both the education content and the emotional support provided by the program when considering program adaptations.

The HeartHealth program was implemented with high fidelity and low participant attrition rates. An important factor for the successful implementation was due to the simplicity of participation and wide acceptance of the program. The acceptance and usability of SMS text messaging programs have previously been demonstrated in other cohorts with CVD [11], as well as cohorts with mental health [34,35] and renal diseases [36] and diabetes [37]. Furthermore, HeartHealth staff reported that the successful implementation was partly enabled through an appropriate, easy-to-use program design, personalized patient contact upon program invitation, and skilled, adaptive staff members. For example, an adaptation initiated by staff members was to contact patients via a phone call to assist with enrollment if they had not responded to the SMS text messaging invitation, resulting in a large increase in enrollment. Previous studies have found self-enrollment to be a barrier for some patients, thereby highlighting the importance of providing assistance with enrollment or offering alternative enrollment modalities [38,39]. It is important to note that despite the HeartHealth program being implemented as a hospital service, the majority of the staff involved were from The University of Sydney. Lower engagement of local clinicians was identified as a barrier. This is consistent with previous observational studies of the implementation of new clinical services, finding that insufficient clinician time [40], lack of clinician motivation [41,42], high staff turnover [43], lack of continuing education [40,42], and an unsupportive organizational culture [40,41] were all barriers to implementation.

Participants identified that a lack of personalization of both message content and delivery was a main program barrier. A common theme from participants was that “one size does not fit all,” with the information not always being relevant and the message timing and modality not suiting all participants. The need for content personalization has been described previously in mHealth interventions among populations with CVD [44,45], and studies comparing personalized with nonpersonalized content on clinical outcomes in cardiovascular populations are lacking [46]; however, benefits have been shown in smoking populations [47]. Advances in machine learning and artificial intelligence will enable the development of personalization of content and responses to patient questions [48]; future research into the design and implementation evaluation of such programs is required.

To enable program sustainability and ongoing improvements, adaptations are required. HeartHealth staff and participants highlighted that the program could be further supported by incorporating a communication channel with health care members to provide information and support when required. It has been shown that patients may feel overwhelmed and unguided on where to find trusted information with the rising tide of health misinformation [49]. Patients with CVD have previously reported that they want their doctor or nurse to recommend information sources [50]. Future research should assess the feasibility and impact on clinical outcomes of an interactive SMS text messaging program before incorporating it into the HeartHealth program.

An important consideration with wider dissemination is the required cost and resources to implement the program. While a comprehensive cost-effectiveness analysis of the HeartHealth program is required, previous community SMS text messaging trials, using the TextCare platform, were found to be cost-saving and health improving in cohorts with CVD [51] and diabetes [52]. Other SMS text messaging programs have demonstrated program cost-effectiveness in patients with renal disease [53] and in smoking cessation campaigns [54]. The cost of the HeartHealth program was significantly less than that of traditional cardiac rehabilitation programs, which has previously been reported to cost between US $631 and US $1457 per participant, depending on the program setting (hospital, home,

XSL•FO
RenderX

or remote) [55-57]. While the HeartHealth program is not designed as a replacement for traditional cardiac rehabilitation programs, it may serve as an effective adjunct. Overall, the HeartHealth program is likely to be appealing to other departments, given the frugal nature of the program, participant benefits, and low-resource requirements needed to implement the program.

Limitations

There were limitations to this study. First, the postintervention survey had a moderate response rate (3533/7964, 44.4%); therefore, feedback may not be representative of all participants. This response rate is in keeping with previous online surveys and may be attributable to survey fatigue during and after the COVID-19 pandemic [58,59]. Second, we did not capture information from those participants who declined the initial SMS text messaging invitation, nor those who opted out of the program after enrollment; therefore, limiting our interpretation of the program's reach and translational ability. Given the program having high participant acceptability and low-resource requirement to implement the program, evaluating long-term program maintenance, awareness, and funding sources will be pertinent for wider dissemination. Third, program effectiveness was assessed using patient-reported outcomes obtained only from a self-selected subgroup of program completers; therefore, the effectiveness results may overstate the true impact of the program due to this inherent bias. However, our previous research underpinning the HeartHealth program conducted using a randomized controlled design demonstrated improved lifestyle behaviors and cardiovascular risk factor profiles [11,16]. To assess program effectiveness more robustly, it would be beneficial to conduct a randomized controlled trial assessing the impact of the program on health care service usage. Fourth, as this was an evaluation of a program implemented into real-world clinical practice, we were limited in data collection and approval to extract linked medical record data. We were able to provide comorbidity data for HeartHealth participants and nonenrollee patients only for the period between April 2020 and April 2022. This limits the ability to describe and characterize those patients who opted in to the program and identify barriers to wider program adoption. Finally, the comorbidity data were drawn from both patient-reported information ("Fidelity and Attrition" section of the "Results" section) and linked medical record data diagnoses (Multimedia Appendix 4), with the latter potentially underrecording conditions such as hypercholesterolemia, hypertension, and diabetes mellitus, as these are typically diagnosed and managed in primary care settings. These differing data sources likely explain the lower prevalence discrepancy of hypercholesterolemia, hypertension, and diabetes mellitus in the "Results" section and in Multimedia Appendix 4.

Implications

The adoption of mobile health technologies has significantly risen over recent years, which can largely be attributed to the COVID-19 pandemic, a heightened focus on telehealth, and the increasing burden on health care services [60]. SMS text messaging programs within cardiovascular populations have been shown to be an effective modality to improve cardiovascular health; however, to date, there have been few that have proceeded to large-scale implementation and a formal scientific evaluation. This study demonstrates the feasibility and use of implementing personalized postdischarge support at low cost. It also identifies important enablers and barriers to implementation. Future scale-up should consider further customization of programs to individuals and broadening availability through personalized language and health literacy.

Conclusions

This thorough implementation evaluation highlights the successful implementation of the HeartHealth program. Participant attrition and perceived lifestyle benefits demonstrate the program effectiveness; additionally, staff and participant feedback has highlighted key program barriers and enablers. These insights provide key learnings for future scale-up and improvement of HeartHealth postdischarge digital support.

Acknowledgments

The authors would like to acknowledge the Westmead Hospital cardiologists, Professor Robert Denniss, Associate Professor Stuart Thomas, Dr Dylan Wynne, Dr Jay Thakkar, Professor Liza Thomas, Dr Mikhail Altman, Associate Professor Eddy Kizana, Dr Mark Cooper, and Dr David Makarious for their role in implementing the HeartHealth program in clinics and inpatient settings.

Funding

The HeartHealth program was funded by the Western Sydney Local Health District. This was supplemented and analysis was supported by an National Health and Medical Research Council investigator program grant awarded to Professor Clara Chow (grant APP1195326).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

BS contributed to conceptualization, data curation, formal analysis, methodology, and writing—original draft. LL and RT participated in data curation, formal analysis, and writing—review and editing. TS contributed to conceptualization, methodology, and writing—review and editing. GS, JC, AT, SZ, and PQ participated in data curation and writing—review and editing. ABI

XSL•FO
RenderX

participated in data curation, project administration, and writing—review and editing. CKC participated in conceptualization, data curation, methodology, supervision, and writing—original draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of text messages used in the HeartHealth program.
[DOCX File, 15 KB - cardio_v10i1e68896_app1.docx]

Multimedia Appendix 2
HeartHealth postintervention participant survey.
[DOCX File, 36 KB - cardio_v10i1e68896_app2.docx]

Multimedia Appendix 3
HeartHealth staff focus group discussion topics.
[DOCX File, 21 KB - cardio_v10i1e68896_app3.docx]

Multimedia Appendix 4
HeartHealth participant and nonenrollee participant cardiovascular comorbidities and English as preferred language data between April 2020 and April 2022 (adapted with permission from Sheahen et al [19]).
[DOCX File, 16 KB - cardio_v10i1e68896_app4.docx]

Multimedia Appendix 5
Roles and duties of HeartHealth staff.
[DOCX File, 16 KB - cardio_v10i1e68896_app5.docx]

Multimedia Appendix 6
Costs associated with the HeartHealth program over the 4-year analysis period, with breakdown of costs during the initial text messaging–only recruitment period during the first 18 months of the program (April 2020 to October 2021) and the subsequent text messaging and phone call recruitment period used during the following 30 months (October 2021 to April 2024).
[DOCX File, 17 KB - cardio_v10i1e68896_app6.docx]

Multimedia Appendix 7
Participant barriers for implementing the HeartHealth program.
[DOCX File, 16 KB - cardio_v10i1e68896_app7.docx]

Multimedia Appendix 8
HeartHealth staff enablers for implementing the HeartHealth program.
[DOCX File, 16 KB - cardio_v10i1e68896_app8.docx]

Multimedia Appendix 9
HeartHealth staff barriers for implementing the HeartHealth program.
[DOCX File, 16 KB - cardio_v10i1e68896_app9.docx]

Multimedia Appendix 10
HeartHealth staff recommended future adaptations for implementing the HeartHealth program.
[DOCX File, 17 KB - cardio_v10i1e68896_app10.docx]

Multimedia Appendix 11
Organization adaptations made to implementing the Heart Health program.
[DOCX File, 16 KB - cardio_v10i1e68896_app11.docx]

References

1. Therapeutic patient education: an introductory guide. World Health Organization. 2023. URL: https://www.who.int/europe/publications/i/item/9789289060219 [accessed 2026-02-04]

XSL•FO
RenderX

2. Improving patient education: a new guide for policy-makers and health professionals to support self-management of chronic conditions. World Health Organization. 2023. URL: https://www.who.int/europe/news/item/14-11-2023-improving-patient-education–a-new-guide-for-policy-makers-and-health-professionals-to-support-self-management-of-chronic-conditions [accessed 2026-02-04]
3. Warsi A, Wang PS, LaValley MP, Avorn J, Solomon DH. Self-management education programs in chronic disease: a systematic review and methodological critique of the literature. Arch Intern Med 2004;164(15):1641-1649. [doi: 10.1001/archinte.164.15.1641] [Medline: 15302634]
4. Reducing risk in heart disease: an expert guide to clinical practice for secondary prevention of coronary heart disease. Cardiac Society of Australia and New Zealand. 2012. URL: https://www.csanz.edu.au/wp-content/uploads/2014/12/2012_HF_CSANZ_Reducing_Risk_in_Heart_Disease.pdf [accessed 2026-02-04]
5. Kwan T, Chua B, Pires D, Feng O, Edmiston N, Longman J. A qualitative analysis of the barriers and enablers faced by Australian rural general practitioners in the non-pharmacological management of congestive heart failure in community dwelling patients. BMC Health Serv Res 2022 Jan 2;22(1):5. [doi: 10.1186/s12913-021-07383-3] [Medline: 34974834]
6. Dhaliwal KK, King-Shier K, Manns BJ, Hemmelgarn BR, Stone JA, Campbell DJT. Exploring the impact of financial barriers on secondary prevention of heart disease. BMC Cardiovasc Disord 2017 Feb 14;17(1):61. [doi: 10.1186/s12872-017-0495-4] [Medline: 28196528]
7. Sérvio TC, Britto RR, de Melo Ghisi GL, et al. Barriers to cardiac rehabilitation delivery in a low-resource setting from the perspective of healthcare administrators, rehabilitation providers, and cardiac patients. BMC Health Serv Res 2019 Sep 2;19(1):615. [doi: 10.1186/s12913-019-4463-9] [Medline: 31477103]
8. Kinsman L, Tham R, Symons J, Jones M, Campbell S, Allenby A. Prevention of cardiovascular disease in rural Australian primary care: an exploratory study of the perspectives of clinicians and high-risk men. Aust J Prim Health 2016 Feb;22(6):510-516. [doi: 10.1071/PY15091] [Medline: 27118176]
9. Rogers HL, Pablo Hernando S, Núñez-Fernández S, et al. Barriers and facilitators in the implementation of an evidence-based health promotion intervention in a primary care setting: a qualitative study. J Health Organ Manag 2021 Sep 2;ahead of print(ahead of print):349-367. [doi: 10.1108/JHOM-12-2020-0512] [Medline: 34464035]
10. Mobile services and coverage. Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts. 2024. URL: https://www.infrastructure.gov.au/media-technology-communications/phone/mobile-services-coverage [accessed 2026-02-04]
11. Chow CK, Redfern J, Hillis GS, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
12. Tam HL, Wong EML, Cheung K, Chung SF. Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic review of randomized controlled trials. JMIR Mhealth Uhealth 2021 Sep 22;9(9):e24527. [doi: 10.2196/24527] [Medline: 34550078]
13. Skinner R, Gonet V, Currie S, Hoddinott P, Dombrowski SU. A systematic review with meta-analyses of text message-delivered behaviour change interventions for weight loss and weight loss maintenance. Obes Rev 2020 Jun;21(6):e12999. [doi: 10.1111/obr.12999] [Medline: 32043809]
14. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4(4):CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
15. Ross ES, Sakakibara BM, Mackay MH, et al. The use of SMS text messaging to improve the hospital-to-community transition in patients with acute coronary syndrome (Txt2Prevent): results from a pilot randomized controlled trial. JMIR Mhealth Uhealth 2021 May 14;9(5):e24530. [doi: 10.2196/24530] [Medline: 33988519]
16. Klimis H, Thiagalingam A, McIntyre D, Marschner S, Von Huben A, Chow CK. Text messages for primary prevention of cardiovascular disease: the TextMe2 randomized clinical trial. Am Heart J 2021 Dec;242:33-44. [doi: 10.1016/j.ahj.2021.08.009] [Medline: 34428440]
17. Redfern J, Thiagalingam A, Jan S, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
18. Klimis H, Thiagalingam A, Chow CK. Text messages for primary prevention of cardiovascular disease: the TextMe2 randomised controlled trial protocol. BMJ Open 2020 Apr 26;10(4):e036767. [doi: 10.1136/bmjopen-2020-036767] [Medline: 32341047]
19. Sheahen B, Marschner S, Min H, et al. Implementation and impact of postdischarge support of cardiovascular patients using text messages the HeartHealth program. JACC Adv 2025 Sep;4(9):102017. [doi: 10.1016/j.jacadv.2025.102017] [Medline: 40773808]
20. Drisko JW, Maschi T. Content Analysis: Oxford University Press; 2016. [doi: 10.1093/acprof:oso/9780190215491.001.0001]
21. Bressman E, Long JA, Honig K, et al. Evaluation of an automated text message-based program to reduce use of acute health care resources after hospital discharge. JAMA Netw Open 2022 Oct 3;5(10):e2238293. [doi: 10.1001/jamanetworkopen.2022.38293] [Medline: 36287564]

XSL•FO
RenderX

22. Clarke M. Impact of an automated text and phone call postdischarge follow-up program on patient satisfaction scores and 30-day hospital readmission among adult patients. : George Washington University; 2023 URL: https://hsrc.himmelfarb.gwu.edu/son_dnp/123 [accessed 2024-09-24]
23. Adusumalli S, Jolly EG, Chokshi NP, Washington LA, Suchoicki L, Walsh MM, et al. Abstract 19699: a change in cardiac rehabilitation referral defaults from opt-in to opt-out increases referral rates among patients with ischemic heart disease. Circulation 2017;136(suppl 1). [doi: 10.1161/circ.136.suppl_1.19699]
24. Adusumalli S, Jolly E, Chokshi NP, et al. Referral rates for cardiac rehabilitation among eligible inpatients after implementation of a default opt-out decision pathway in the electronic medical record. JAMA Netw Open 2021 Jan 4;4(1):e2033472. [doi: 10.1001/jamanetworkopen.2020.33472]
25. Etheridge C, Sinyard RD, Brindle ME. Chapter 90—implementation research. In: Osband AJ, Newell PC, Bakal JA, Eltorai AEM, editors. Translational Surgery: Academic Press; 2023:563-573. [doi: 10.1016/B978-0-323-90300-4.00043-4]
26. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
27. Mayo NL, Russell HA, Holt K, Williams GC. Implementation of a self-determination based clinical program to reduce cardiovascular disease risk. J Health Psychol 2022 Nov;27(13):2898-2908. [doi: 10.1177/13591053211068124] [Medline: 35086378]
28. Tzelepis F, Sanson-Fisher RW, Zucca AC, Fradgley EA. Measuring the quality of patient-centered care: why patient-reported measures are critical to reliable assessment. Patient Prefer Adherence 2015;9:831-835. [doi: 10.2147/PPA.S81975] [Medline: 26150703]
29. Bradshaw J, Siddiqui N, Greenfield D, Sharma A. Kindness, listening, and connection: patient and clinician key requirements for emotional support in chronic and complex care. J Patient Exp 2022;9:23743735221092627. [doi: 10.1177/23743735221092627] [Medline: 35434291]
30. Han E, Shiraz F, Haldane V, et al. Biopsychosocial experiences and coping strategies of elderly ESRD patients: a qualitative study to inform the development of more holistic and person-centred health services in Singapore. BMC Public Health 2019 Aug 14;19(1):1107. [doi: 10.1186/s12889-019-7433-6] [Medline: 31412824]
31. Allen D, Scarinci N, Hickson L. The nature of patient- and family-centred care for young adults living with chronic disease and their family members: a systematic review. Int J Integr Care 2018 May 18;18(2):14. [doi: 10.5334/ijic.3110] [Medline: 30127698]
32. Rathert C, Williams ES, McCaughey D, Ishqaidef G. Patient perceptions of patient-centred care: empirical test of a theoretical model. Health Expect 2015 Apr;18(2):199-209. [doi: 10.1111/hex.12020] [Medline: 23176054]
33. Singleton AC, Raeside R, Hyun KK, et al. A National Health and Wellness SMS text message program for breast cancer survivors during COVID-19 (EMPOWER-SMS COVID-19): mixed methods evaluation using the RE-AIM framework. J Med Internet Res 2023 Jul 25;25:e45164. [doi: 10.2196/45164] [Medline: 37490319]
34. Patel S, Akhtar A, Malins S, et al. The acceptability and usability of digital health interventions for adults with depression, anxiety, and somatoform disorders: qualitative systematic review and meta-synthesis. J Med Internet Res 2020 Jul 6;22(7):e16228. [doi: 10.2196/16228] [Medline: 32628116]
35. Shariful Islam SM, Chow CK, Redfern J, et al. Effect of TEXT messaging on depression in patients with coronary heart disease: a substudy analysis from the TEXT ME randomised controlled trial. BMJ Open 2019 Feb 20;9(2):e022637. [doi: 10.1136/bmjopen-2018-022637] [Medline: 30787075]
36. Dawson J, Campbell KL, Craig JC, et al. A text messaging intervention for dietary behaviors for people receiving maintenance hemodialysis: a feasibility study of KIDNEYTEXT. Am J Kidney Dis 2021 Jul;78(1):85-95. [doi: 10.1053/j.ajkd.2020.11.015] [Medline: 33421456]
37. Middleton T, Constantino M, McGill M, et al. An enhanced SMS text message-based support and reminder program for young adults with type 2 diabetes (TEXT2U): randomized controlled trial. J Med Internet Res 2021 Oct 21;23(10):e27263. [doi: 10.2196/27263] [Medline: 34524102]
38. Speirs KE, Grutzmacher SK, Munger AL, Messina LA. Recruitment and retention in an SMS-based health education program: lessons learned from Text2BHealthy. Health Informatics J 2016 Sep;22(3):651-658. [doi: 10.1177/1460458215577995] [Medline: 25916832]
39. Gazmararian JA, Elon L, Yang B, Graham M, Parker R. Text4baby program: an opportunity to reach underserved pregnant and postpartum women? Matern Child Health J 2014 Jan;18(1):223-232. [doi: 10.1007/s10995-013-1258-1] [Medline: 23494485]
40. Wallis L. Barriers to implementing evidence-based practice remain high for U.S. nurses. Am J Nurs 2012;112(12):15. [doi: 10.1097/01.NAJ.0000423491.98489.70]
41. Ramón C, Nievas-Soriano BJ, García-González J, Alarcón-Rodríguez R, Requena-Mullor M, Lozano-Paniagua D. Motivation and barriers to research among nursing professionals in Southeast Spain. Healthcare (Basel) 2022 Apr 2;10(4):675. [doi: 10.3390/healthcare10040675] [Medline: 35455852]
42. Curtis K, Fry M, Shaban RZ, Considine J. Translating research findings to clinical nursing practice. J Clin Nurs 2017 Mar;26(5-6):862-872. [doi: 10.1111/jocn.13586] [Medline: 27649522]

XSL•FO
RenderX

43. Baron AN, Hemler JR, Sweeney SM, et al. Effects of practice turnover on primary care quality improvement Implementation. Am J Med Qual 2020 Jan;35(1):16-22. [doi: 10.1177/1062860619844001]
44. Atluri N, Mishra SR, Anderson T, et al. Acceptability of a text message‐based mobile health intervention to promote physical activity in cardiac rehabilitation enrollees: a qualitative substudy of participant perspectives. J Am Heart Assoc 2024 Jan 16;13(2):e030807. [doi: 10.1161/JAHA.123.030807]
45. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiolog 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913]
46. Shariful Islam SM, Farmer AJ, Bobrow K, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019;6(2):e001017. [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
47. Lin H, Li M, Xiao L, Chang C, Liu GG. Efficacy of personalised text message intervention in reducing smoking frequency and amount for non-abstinent smokers: a double-blind, randomised controlled trial. J Glob Health 2023 Oct 27;13:04133. [doi: 10.7189/jogh.13.04133] [Medline: 37883205]
48. Farmer A, French D, Bartlett K. Chapter 14—short message service (SMS) text messages in health care. In: Klonoff DC, Kerr D, Espinoza JC, editors. Diabetes Digital Health, Telehealth, and Artificial Intelligence: Academic Press; 2024:171-183 URL: https://doi.org/10.1016/B978-0-443-13244-5.00020-1 [accessed 2024-09-30]
49. Daraz L, Morrow AS, Ponce OJ, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891. [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
50. Wells S, Mahony F, Lee A, et al. Preferred format and strategies for seeking and trusting online health information: a survey of cardiology outpatient attendees across three New Zealand hospitals. J Prim Health Care 2024 Sep;16(3):270-277. [doi: 10.1071/HC23143] [Medline: 39321076]
51. Burn E, Nghiem S, Jan S, et al. Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. Heart 2017 Jun;103(12):893-894. [doi: 10.1136/heartjnl-2016-310195] [Medline: 28235776]
52. Islam SMS, Peiffer R, Chow CK, et al. Cost-effectiveness of a mobile-phone text messaging intervention on type 2 diabetes—a randomized-controlled trial. Health Policy Technol 2020 Mar;9(1):79-85. [doi: 10.1016/j.hlpt.2019.12.003]
53. Dawson J, Howell M, Howard K, et al. Cost-effectiveness of a mobile phone text messaging program (KIDNEYTEXT) targeting dietary behaviours in people receiving haemodialysis. J Hum Nutr Diet 2022 Oct;35(5):765-773. [doi: 10.1111/jhn.12937] [Medline: 34323334]
54. Cobos-Campos R, Mar J, Apiñaniz A, et al. Cost-effectiveness analysis of text messaging to support health advice for smoking cessation. Cost Eff Resour Alloc 2021 Feb 15;19(1):9. [doi: 10.1186/s12962-021-00262-y] [Medline: 33588885]
55. Edwards K, Jones N, Newton J, et al. The cost-effectiveness of exercise-based cardiac rehabilitation: a systematic review of the characteristics and methodological quality of published literature. Health Econ Rev 2017 Oct 19;7(1):37. [doi: 10.1186/s13561-017-0173-3] [Medline: 29052044]
56. Oehler AC, Holmstrand EC, Zhou L, et al. Cost analysis of remote cardiac rehabilitation compared with facility-based cardiac rehabilitation for coronary artery disease. Am J Cardiol 2024 Jan 1;210:266-272. [doi: 10.1016/j.amjcard.2023.08.061] [Medline: 37973439]
57. Southard BH, Southard DR, Nuckolls J. Clinical trial of an Internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
58. de Koning R, Egiz A, Kotecha J, et al. Survey fatigue during the COVID-19 pandemic: an analysis of neurosurgery survey response rates. Front Surg 2021;8:690680. [doi: 10.3389/fsurg.2021.690680]
59. Krieger N, LeBlanc M, Waterman PD, Reisner SL, Testa C, Chen JT. Decreasing survey response rates in the time of COVID-19: implications for analyses of population health and health inequities. Am J Public Health 2023 Jun;113(6):667-670. [doi: 10.2105/AJPH.2023.307267] [Medline: 37023386]
60. Admitted patient care. Australian Institute of Health and Welfare. 2023. URL: https://www.aihw.gov.au/reports-data/myhospitals/sectors/admitted-patients [accessed 2026-02-04]

Abbreviations

CVD: cardiovascular disease
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

Edited by A Coristine; submitted 17.Nov.2024; peer-reviewed by J Salinas, M Carrington, MDD Smith; revised version received 21.Jan.2026; accepted 22.Jan.2026; published 11.Mar.2026.

Please cite as:
Sheahen B, Laranjo L, Trivedi R, Shaw T, Sivagangabalan G, Chong J, Thiagalingam A, Zaman S, Qian P, Indrawansa AB, Chow CK
The HeartHealth Program: A Mixed Methods Study of a Community-Based Text Messaging Support Program for Patients With Cardiovascular Disease From 2020 to 2024
JMIR Cardio 2026;10:e68896
URL: https://cardio.jmir.org/2026/1/e68896
doi:10.2196/68896

© Brodie Sheahen, Liliana Laranjo, Ritu Trivedi, Tim Shaw, Gopal Sivagangabalan, James Chong, Aravinda Thiagalingam, Sarah Zaman, Pierre Qian, Anupama Balasuriya Indrawansa, Clara Kayei Chow. Originally published in JMIR Cardio (https://cardio.jmir.org), 11.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Human vs Artificial Intelligence in Detecting Coronary Occlusions in ST-Elevation Myocardial Infarction: Retrospective Study

Aurelia Zimmerli[1*], MD; Marion Dupre[1*], MD; Quentin Liabot[1], MD; C-Yasmine Sellah[1]; David Meier[1], MD; Adil Salihu[1], MD; Victor Weerts[1], MD; Vladimir Rubimbura[1], MD; Olivier Roux[2], MD; Olivier Muller[1], MD, PhD; Olivier Hugli[3], MD, MPH; Pierre-Nicolas Carron[3], MD; Stephane Fournier[1], MD, PhD

[1]Department of Cardiology, University Hospital of Lausanne, Rue du Bugnon 46, Lausanne, Switzerland

[2]Hirslanden Klinik Im Park, Zurich, Switzerland

[3]Emergency Department, University Hospital of Lausanne, Lausanne, Switzerland

[*]these authors contributed equally

Corresponding Author:
Stephane Fournier, MD, PhD
Department of Cardiology, University Hospital of Lausanne, Rue du Bugnon 46, Lausanne, Switzerland

Abstract

Background: Occlusion myocardial infarction (OMI) on the electrocardiogram (ECG) can present as ST-elevation myocardial infarction (STEMI), non-STEMI, ST depression, or other patterns (eg, de Winter T waves and hyperacute T waves), posing a diagnostic challenge. Even experienced physicians may overlook OMI on ECG, leading to delays in catheterization laboratory activation. Recently, smartphone apps using AI have been shown to exhibit promising accuracy in ECG interpretation in large datasets.

Objective: The objective of our study was to retrospectively compare the diagnostic performance of an AI-based application (PMCardio) with that of interventional cardiologists in interpreting nonpathognomonic ECGs that generated debates about diagnosis.

Methods: In total, 33 ECGs that were diagnostically challenging upon patient admission to the emergency department were included. A group of 3 interventional cardiologists, blinded to the final diagnosis, independently assessed whether the ECG suggested OMI requiring immediate invasive coronary angiography. The diagnostic performance of the cardiologists was compared to that of the AI application, using the results of coronary angiography as the gold standard and considering any lesion with ≥90% stenosis as positive.

Results: Of the 33 patients, 29 (88%) had a final diagnosis of cardiac origin for their chest pain, including 23 acute coronary syndromes, of which 13 (57%) were OMI. The 3 cardiologists achieved sensitivities of 62%, 85%, and 62%; negative predictive values (NPVs) of 74%, 83%, and 69%; specificities of 70%, 50%, and 55%; and positive predictive values (PPVs) of 57%, 52%, and 47%. Their overall misclassification rates were 33%, 36%, and 42%. In contrast, AI analysis demonstrated a sensitivity and NPV of 100%, a specificity of 60%, a PPV of 62%, and an overall misclassification rate of 24%.

Conclusions: In this retrospective study of challenging ECGs, an AI application demonstrated better diagnostic performance of OMI as compared to cardiologists. These findings highlight the potential of AI to enhance diagnostic accuracy and support clinical decision-making.

(JMIR Cardio 2026;10:e88399) doi:10.2196/88399

KEYWORDS

artificial intelligence; AI; coronary artery disease; CAD; myocardial infarction; electrocardiogram; ECG

Introduction

In patients presenting with chest pain, electrocardiogram (ECG) allows one to distinguish between ST-elevation myocardial infarction (STEMI) and non-STEMI (NSTEMI). STEMI is generally caused by complete occlusion of a coronary artery, leading to ST-segment elevation on the ECG and requires immediate coronary angiography. Conversely, NSTEMI is typically associated with subtotal occlusion, resulting in ischemia-reperfusion phenomenon without ST-segment elevation. In this context, guidelines generally recommend coronary angiography to be performed within 24 hours [1].

However, this distinction has important limitations. Indeed, while STEMI criteria are highly specific (>95%), their sensitivity is much lower, at just above 40%, corresponding to a positive likelihood ratio of 12.5 but a negative likelihood ratio of only 0.58 [2]. On the other hand, more than 25% of NSTEMI

XSL•FO
RenderX

patients are found to have a totally occluded culprit during coronary angiography [3]. Reliance on STEMI criteria alone to detect occlusion myocardial infarction (OMI) criteria is thus problematic, and a significant proportion of patients might receive delayed reperfusion, leading to larger infarcts and higher mortality rates.

OMI can present with different ECG patterns, including STEMI and NSTEMI, as well as atypical patterns sometimes referred to as "STEMI equivalents" [4,5]. Furthermore, new patterns are regularly described [6], adding to the diagnostic challenge.

Rapid and accurate diagnosis requires expertise in ECG interpretation in emergency settings, an expertise that is not always widely available depending on the timing (off-hours) and setting (primary care centers) of presentation. AI-based applications have been developed to aid ECG analysis, and recent studies have shown that AI can outperform both emergency physicians and interventional cardiologists in detecting coronary occlusion [7,8].

At Lausanne University Hospital (CHUV), ECGs of patients considered candidates for immediate reperfusion are transmitted from the prehospital setting or other hospital emergency departments (EDs) to a senior emergency physician, who decides whether urgent coronary angiography is warranted. When the ECG does not meet STEMI or STEMI-equivalent criteria, the on-call interventional cardiologist is consulted. This study aims to compare the diagnostic performance of an AI-based application (PMCardio) with that of interventional cardiologists in interpreting a consecutive series of challenging ECGs that do not meet STEMI criteria.

Methods

Data Collection

This is a single-center retrospective study conducted at CHUV. It represents the extension of a master's thesis initially conducted as a quality improvement project. The original aim was to assess the use of a generic email system designed to dispatch ECGs of undetermined significance to the on-call interventional cardiologist. As this specific cohort was composed exclusively of ECGs for which a diagnostic uncertainty was present and whose final diagnosis had been adjudicated, it was used to evaluate the performance of an AI-based ECG interpretation tool.

All consecutive ECGs received by emergency physicians between February 26, 2019, and May 21, 2020, and submitted to the interventional cardiologist through this dispatch system were included. Only ECGs from patients who provided written informed consent and had a definitive established diagnosis were included in this analysis.

Data Analysis

The ECGs were anonymized and independently evaluated by 3 experienced interventional cardiologists. Each cardiologist was asked to answer the following question: "If this ECG were associated with a compatible clinical presentation, would you activate the STEMI pathway for immediate invasive coronary angiography?" The same ECGs were uploaded as ECG images into the AI-based application PMCardio (PMcardio for Organizations, version 1.7.0; Powerful Medical).

PMCardio is a CE-marked AI medical device [9] that uses a deep neural network designed for the automated interpretation of 12-lead ECG waveforms. The initial version of the model (eOMI version 1) was trained on >18,000 ECGs from patients with chest pain and validated in an international study [7]. Subsequently, an updated version of the model (aOMI version 1) was developed and recalibrated on >200,000 ECGs to improve performance in identifying STEMI and STEMI-equivalent patterns associated with active coronary occlusion requiring emergent reperfusion. Across development iterations, the model was trained on a large multinational dataset comprising >200,000 ECG recordings linked to clinical outcomes and angiographic findings. The reference standard used for model training consisted of angiographically confirmed occlusive myocardial infarction, defined as the identification of a culprit vessel with either total occlusion or critical stenosis associated with acute myocardial infarction and biomarker elevation, as adjudicated using the full clinical record. In the present validation study, ECGs were analyzed using the PMcardio for Organizations smartphone app (version 1.7.0). ECG tracings were uploaded as photographs of printed ECGs captured via smartphone, and the app automatically digitized the ECG images into waveform signals using a validated AI-based ECG digitization pipeline before applying the diagnostic model [10]. Importantly, the datasets used for model development and training came from a separate study and were entirely independent of the validation cohort used in the present study. No overlap or data leakage occurred between the development or training datasets and the validation cohort. The present analysis therefore represents an external evaluation of the algorithm in a separate cohort of diagnostically challenging ECGs, minimizing the risk of overfitting.

The diagnostic concordance of cardiologists and the AI was assessed against the final clinical diagnosis of OMI, with coronary angiography serving as the gold standard and lesions with ≥90% stenosis classified as positive. The angiographic report included assessments of culprit vessel identification and thrombolysis in myocardial infarction (TIMI) flow. Final diagnostic adjudication was corroborated using the discharge summary, which incorporated troponin values and other clinical and imaging data.

Statistical Analysis

All statistical analyses were conducted using SPSS (version 24.0; IBM Corp). Figures were generated using Prism (version 7.0a; GraphPad Software). General patient data are presented as mean (SD) or median (IQR), depending on the results of normality tests. Qualitative analyses were performed to categorize ECG diagnostic uncertainties into distinct groups. The misclassification rate was defined as the proportion of incorrectly classified cases, calculated as the sum of false-positive and false-negative results divided by the total number of cases. Conversely, diagnostic accuracy was defined as the proportion of correctly classified cases, calculated as the sum of true-positive and true-negative results divided by the total number of cases. To evaluate the differences in diagnostic

XSL•FO
RenderX

performance between the AI application and the cardiologists, we used the McNemar test based on the consensus adjudication of the 3 cardiologists rather than individual reader assessments, as this approach reduces interreader variability and better reflects real-world expert evaluation of ambiguous ECGs.

Ethical Considerations

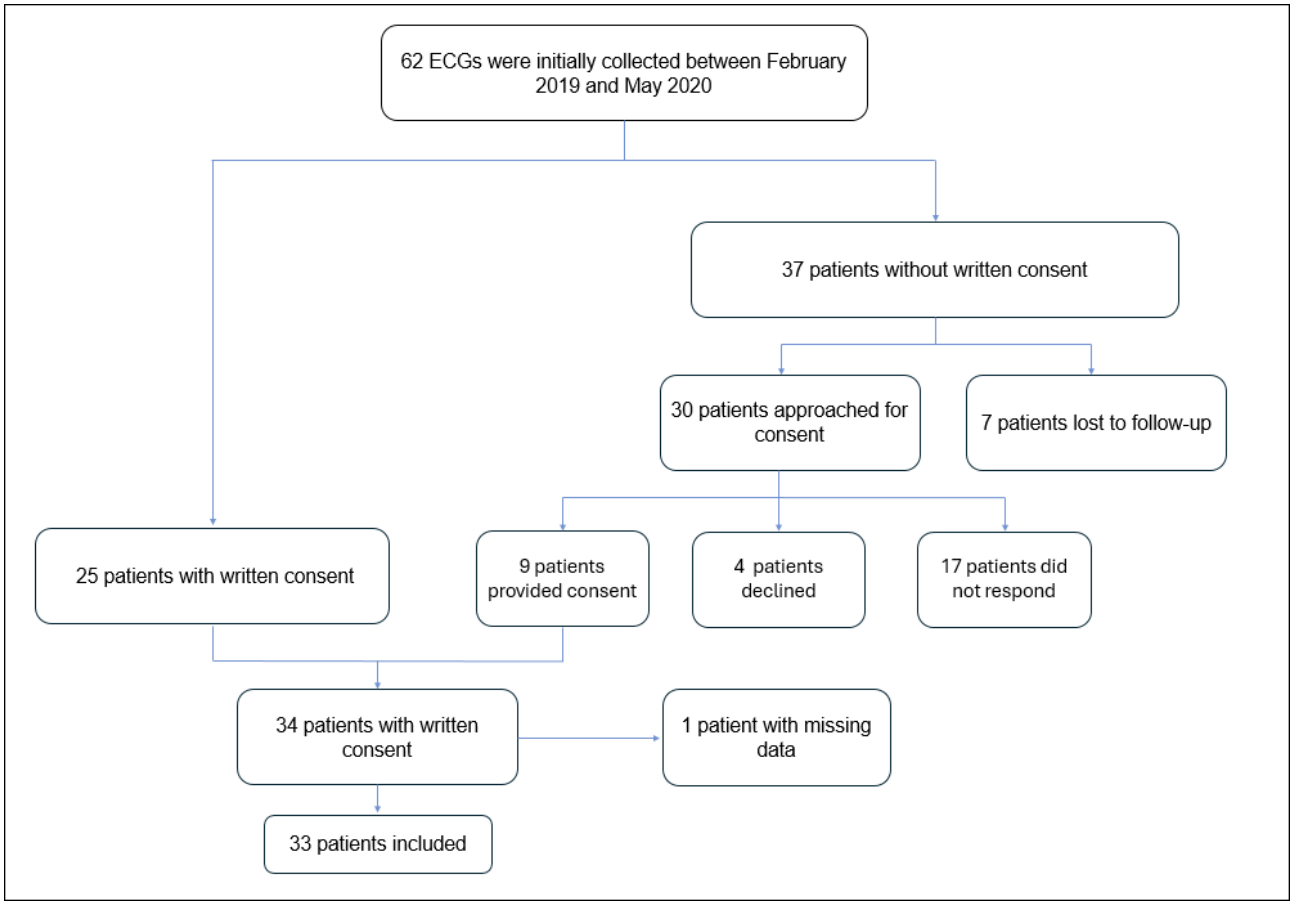
This study received ethics approval from the Cantonal Ethics Committee of Vaud, Switzerland (project ID: 2020-01451). Written informed consent was obtained from all included patients. The consent covered secondary analysis of clinical data. All ECGs were anonymized prior to analysis. No directly identifiable data were accessible to investigators. Participants did not receive any financial compensation for their participation in the study.

Results

Overview

A total of 62 ECGs were initially collected between February 26, 2019, and May 21, 2020. Among them, 25 patients had previously provided general consent for the use of their clinical data for research purposes. For the remaining patients, additional consent was sought by mail when contact information was available. When necessary, referring hospitals were contacted to retrieve missing contact details. Overall, 30 consent forms were sent, resulting in 9 additional patients providing consent, 4 refusals, and 17 nonresponses. Thus, a total of 34 patients consented to inclusion. Final clinical diagnoses were obtained from electronic medical records or by contacting referring institutions when necessary. One patient could not be reached, resulting in 33 patients with complete data included in the final analysis (Figure 1). The study included 21 men (63%) and 12 women (36%), with a mean age of 69 (SD 8.0) years for men and 74 (SD 3.25) years for women.

Figure 1. Flow diagram of patient selection. A total of 62 electrocardiograms (ECGs) were collected between February 2019 and May 2020. After additional consent procedures, 34 patients provided consent. Final diagnoses were obtained from medical records or referring institutions. One patient was excluded due to missing data, resulting in 33 patients included in the final analysis.

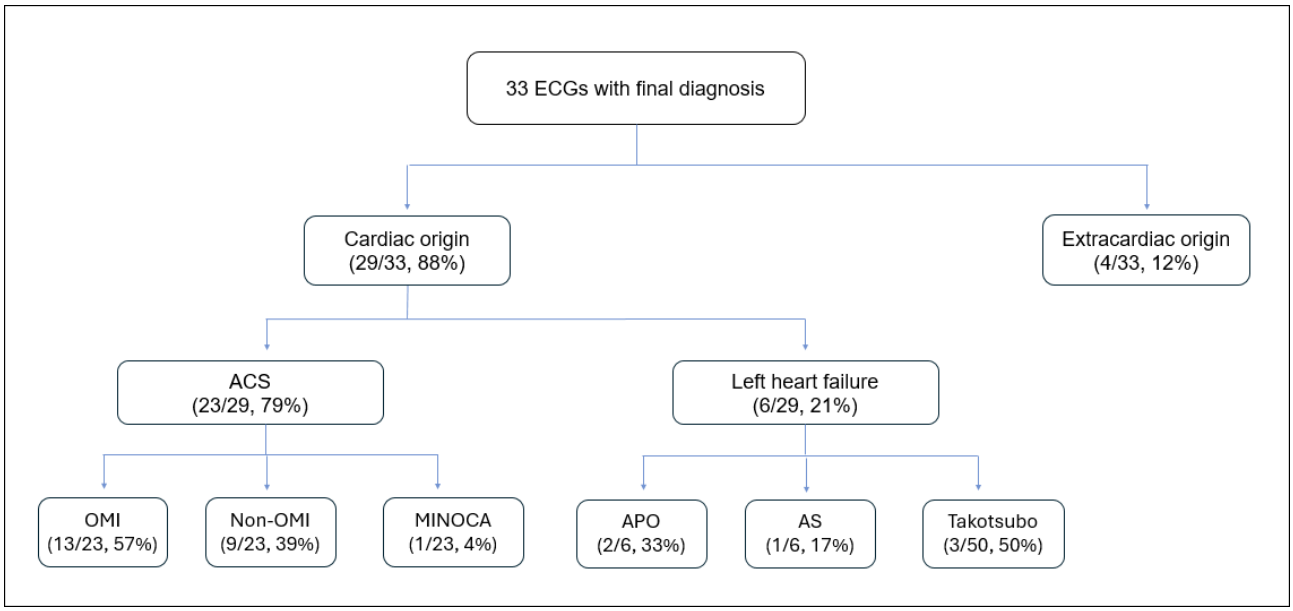


Final Diagnosis

Among 33 patients, 29 (88%) were confirmed to have a cardiac origin of their symptoms, including 23 acute coronary syndromes (ACS), of which 13 (57%) were OMI. The remaining 4 patients were diagnosed with extracardiac conditions (Figure 2).

XSL•FO
RenderX

Figure 2. Flow diagram of the final diagnosis in a retrospective study of 33 patients undergoing electrocardiogram (ECG) evaluation. The study included patients with suspected cardiac events at Lausanne University Hospital between February 2019 and May 2020. ACS: acute coronary syndrome; APO: acute pulmonary oedema; AS: aortic stenosis; MINOCA: myocardial infarction with nonobstructive coronary arteries; OMI: occlusion myocardial infarction.

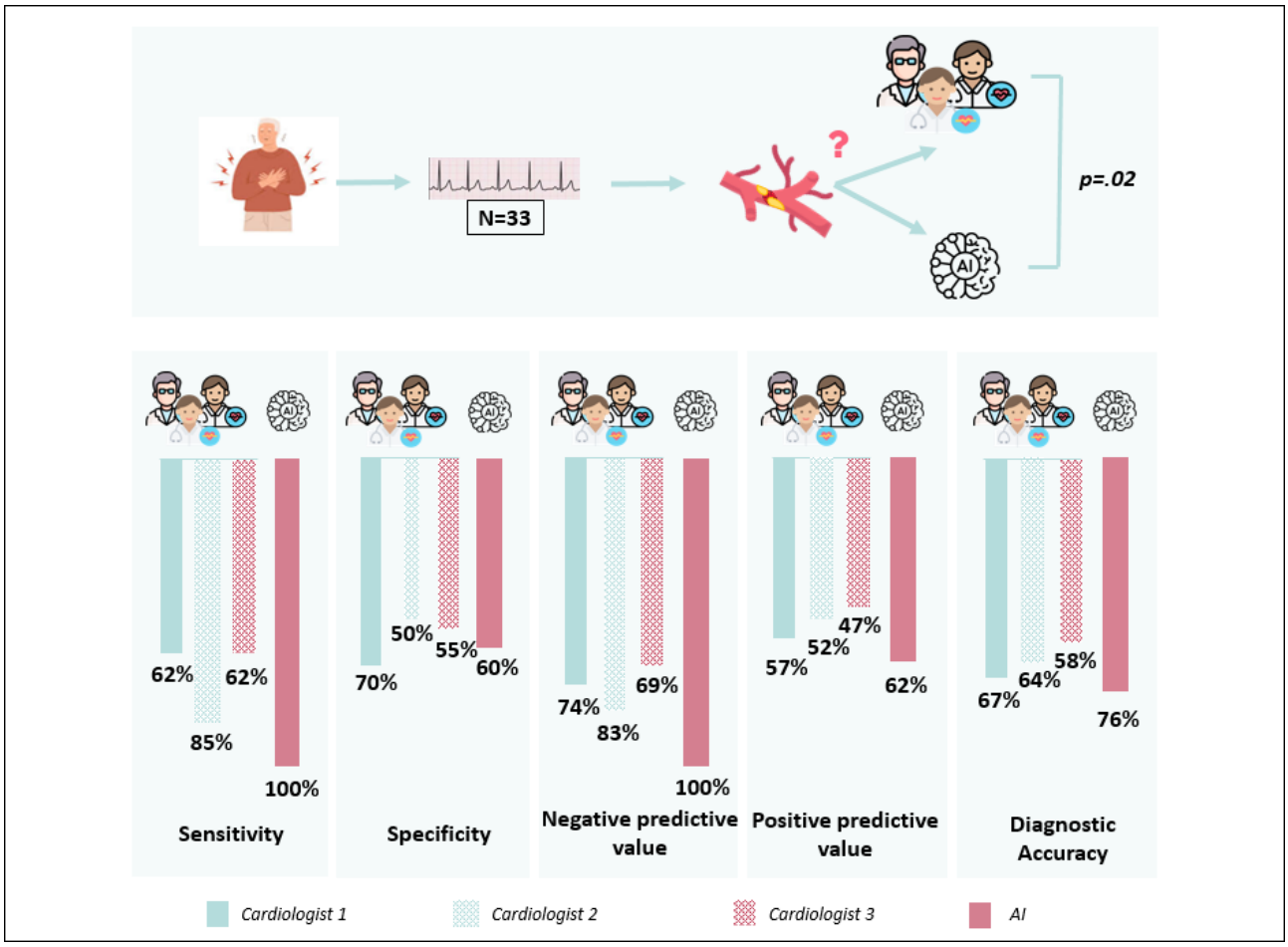


Analysis by Interventional Cardiologists

Cardiologist 1 correctly identified 8 out of 13 patients with an occluded or nearly occluded vessel (≥90% stenosis), corresponding to a sensitivity of 62%. He recommended an emergency procedure in 14 patients, among whom 8 had an occlusion or near-occlusion, yielding a positive predictive value (PPV) of 57%. Among the 20 patients without such lesions, he refrained from indicating an urgent catheterization in 14 cases, resulting in a specificity of 70% (95% CI 50% - 90%). Finally, in the 19 cases where no emergency angiography was advised, 5 patients were later found to have an occluded or nearly occluded vessel, leading to a negative predictive value (NPV) of 74% (95% CI 54% - 94%; Figure 3).

XSL•FO
RenderX

Figure 3. The diagnostic performance of cardiologists vs an AI algorithm for detecting occlusion myocardial infarction (OMI) in a retrospective study of 33 difficult-to-interpret electrocardiograms (ECGs). Using the exact binomial version of the McNemar test, the AI system performed significantly better than the consensus diagnosis of cardiologists (*P*=.02).



Cardiologist 2 correctly identified 11 out of 13 patients with an occluded or nearly occluded vessel (≥90% stenosis), corresponding to a sensitivity of 85% (95% CI 64%-100%). He recommended an emergency procedure in 21 patients, among whom 11 had an occlusion or near-occlusion, yielding a PPV of 52% (95% CI 30%-74%). Among the 20 patients without such lesions, he refrained from indicating an urgent catheterization in 10 cases, resulting in a specificity of 50% (95% CI 34%-76%). Finally, in the 12 cases where no emergency angiography was advised, 2 patients were later found to have an occluded or nearly occluded vessel, leading to an NPV of 83% (95% CI 62%-100%; Figure 3).

Cardiologist 3 correctly identified 8 out of 13 patients with an occluded or nearly occluded vessel (≥90% stenosis), corresponding to a sensitivity of 62% (95% CI 35%-88%). He recommended an emergency procedure in 17 patients, among whom 8 had an occlusion or near-occlusion, yielding a PPV of 47% (95% CI 23%-71%). Among the 20 patients without such lesions, he refrained from indicating an urgent catheterization in 11 cases, resulting in a specificity of 55% (95% CI 34%-76%). Finally, in the 16 cases where no emergency angiography was advised, 5 patients were later found to have an occluded or nearly occluded vessel, leading to an NPV of 69% (95% CI 46%-92%; Figure 3).

When comparing the diagnosis provided by the 3 cardiologists, all 3 agreed on the same diagnosis for 18 patients, while partial agreement (2 of 3 cardiologists) was observed for 15 patients.

Analyses by the AI-Based Application

The AI-based application identified 13 out of 13 patients with an OMI, corresponding to a sensitivity of 100% (95% CI 75%-100%). An emergency procedure was recommended in 21 patients, yielding a PPV of 62% (95% CI 39%-85%). Among the 20 patients without OMI, urgent catheterization was recommended in 8 cases, resulting in a specificity of 60% (95% CI 38%-82%) and an NPV of 100% (95% CI 73%-100%; Figure 3).

Furthermore, AI demonstrated a misclassification rate of 24%, whereas the interventional cardiologists had misclassification rates ranging from 33% to 42%.

To compare human performance, a consensus diagnosis among cardiologists was adjudicated on majority vote in cases of disagreement. This analysis showed that the AI system performed significantly better than the cardiologists (*P*=.02).

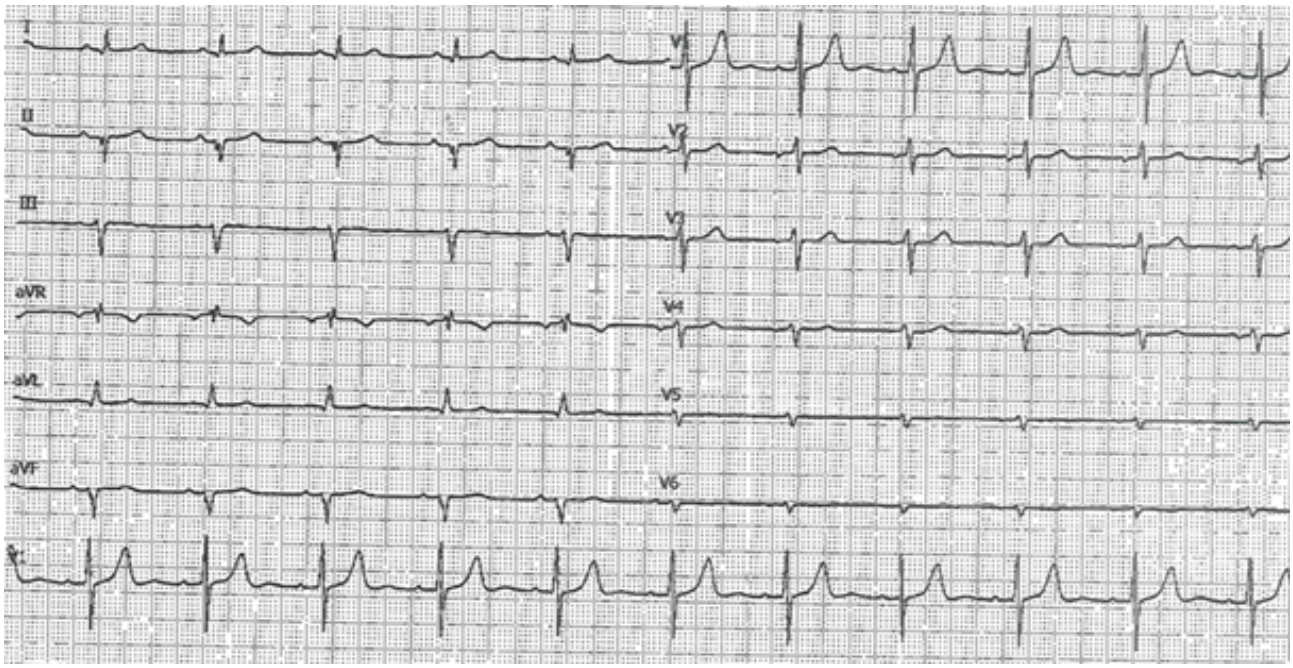
Misclassified Pathological ECG

Among the ECGs analyzed, 1 case was particularly illustrative of the diagnostic divergence between OMI and its ECG manifestations (Figure 4). Coronary angiography subsequently revealed an occlusion of a marginal branch artery. The AI-based

XSL•FO
RenderX

application correctly identified the ECG as an OMI and recommended immediate coronary angiography. Conversely, none of the interventional cardiologists interpreted the ECG as warranting urgent invasive evaluation.

Figure 4. Representative electrocardiogram (ECG) from a patient included in the study cohort with angiographically confirmed occlusion of a marginal branch artery. In this case, none of the 3 cardiologists identified the diagnosis on ECG, whereas the AI algorithm correctly detected occlusion myocardial infarction.



Discussion

Principal Findings

In this retrospective study of 33 real-life challenging ECGs, 13 (39%) were confirmed as OMI by coronary angiography. The sensitivity of the cardiologist for detecting OMI ranged from 62% to 85%, while the AI-based application achieved a sensitivity of 100%. The AI-based application also had a lower misclassification rate (24%) compared to cardiologists (33%-42%), demonstrating a superior diagnostic performance. Notably, the AI application did not miss any OMI cases, a finding of particular clinical relevance given the consequences of delayed reperfusion. This further supports the importance of classifying patients presenting with chest pain as OMI or non-OMI, rather than relying solely on conventional ECG criteria, to guide the timing of coronary revascularization [11].

Clinical Implications

The higher number of false positives observed among cardiologists compared to the AI-based tool should be interpreted with caution. The misclassification rate was used as a global measure of diagnostic error; however, this metric combines false-positive and false-negative results and therefore assumes equal clinical weight between them. In the context of suspected OMI, these errors have different implications. False negatives may lead to delayed reperfusion and potentially worse outcomes, whereas false positives mainly result in unnecessary catheterization laboratory activation and increased resource use. For this reason, sensitivity and NPV remain particularly important metrics when evaluating diagnostic tools intended to support early detection of OMI. In our analysis, any case in which an OMI was suspected where angiogram showed a stenosis <90% was classified as a false positive. However, this binary categorization does not necessarily capture the full clinical rationale underlying the decision to proceed with an invasive strategy. Moreover, some occlusions can spontaneously recanalize, notably by responding to antiplatelet and anticoagulant therapy (eg, aspirin and heparin) administered before coronary angiography. This dynamic can lead to underestimation of the initial severity of the coronary lesion. Taken together, these considerations suggest that the reported false-positive rate, especially among cardiologists, should be interpreted with clinical nuance and not as a strict marker of diagnostic error.

Comparison With Prior Work

This study was conducted in a highly selected cohort of ECGs that prompted real-time consultation with an interventional cardiologist. As such, the prevalence of OMI in our sample (13/33, 39%) is substantially higher than that observed in an unselected ED population. Therefore, our results should be interpreted within the specific context of diagnostically ambiguous ECGs rather than as a reflection of screening performance in all-comers chest pain populations.

AI might play a central role in the analysis of ECGs in patients with suspected ACS [7]. Diagnosis of OMI on the ECG remains a complex challenge, as studies have shown that approximately one-third of OMI cases do not present with a typical ST-segment elevation on the initial ECG [12]. Moreover, only 49% of OMI cases are recognized by cardiologists, leading to undertreatment of a time-sensitive and potentially lethal pathology [13,14]. Our study showed that in this specific population with complex ECGs, AI demonstrated higher sensitivity than cardiologists. Furthermore, cardiologists exhibited significant interindividual

XSL•FO
RenderX

variability, with 18 ECGs receiving the same diagnosis from all 3 specialists. These findings underscore the potential of AI as a valuable tool in ECG interpretation, particularly in detecting subtle or ambiguous signs of OMI. By improving diagnostic accuracy and assisting clinical decision-making, AI has the potential to enhance patient outcomes, making its integration into clinical practice highly relevant. Our study contributes to the growing body of literature supporting the use of AI for OMI detection and provides evidence of its validity in a local clinical setting.

Limitations

This study has several limitations. First, the small sample size is due to the specific population defined by diagnostically challenging ECGs, adjudicated final diagnosis, and written informed consent. As a result, critically ill patients were excluded, which may have introduced survivorship bias. These findings should therefore be validated in a larger and more representative cohort. Second, no clinical data were available for analysis, including patient presentation (symptom onset, duration, and characteristics), which may have impacted the diagnostic performance of the cardiologists and does not fully reflect real-world clinical conditions. Additionally, the quality of some ECGs was suboptimal, with incomplete tracings. Finally, OMI defined by ≥90% stenosis provides an objective angiographic benchmark but may oversimplify clinical reality. Unstable thrombotic lesions with <90% stenosis or reduced TIMI flow can justify urgent reperfusion. Therefore, some false positives may represent appropriate clinical decisions rather than diagnostic errors.

Conclusions

AI-based ECG interpretation showed a higher sensitivity and a lower misclassification rate compared to interventional cardiologists for the detection of OMI in this cohort of challenging ECGs. These findings suggest that AI may represent a valuable adjunctive tool to support clinical decision-making. Further studies in larger and more representative populations are needed to confirm these results and better define the role of AI in emergency cardiac care.

Acknowledgments

ChatGPT (OpenAI) was used solely for grammar and language editing and was not used for content generation.

Funding

The authors received no financial support for the research, authorship, or publication of this paper. The study was not sponsored by PMCardio, and the company had no involvement in the study's conception, analysis, or decision to publish the results.

Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

Authors' Contributions

AZ, MD, and SF designed the study and drafted the manuscript. AZ and MD collected and analyzed the data. All authors critically revised the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Neumann FJ, Sousa-Uva M, Ahlsson A, et al. 2018 ESC/EACTS guidelines on myocardial revascularization. Eur Heart J 2019 Jan 7;40(2):87-165. [doi: 10.1093/eurheartj/ehy394] [Medline: 30165437]
2. de Alencar Neto JN, Scheffer MK, Correia BP, Franchini KG, Felicioni SP, De Marchi MF. Systematic review and meta-analysis of diagnostic test accuracy of ST-segment elevation for acute coronary occlusion. Int J Cardiol 2024 May 1;402:131889. [doi: 10.1016/j.ijcard.2024.131889] [Medline: 38382857]
3. Wang TY, Zhang M, Fu Y, et al. Incidence, distribution, and prognostic impact of occluded culprit arteries among patients with non-ST-elevation acute coronary syndromes undergoing diagnostic angiography. Am Heart J 2009 Apr;157(4):716-723. [doi: 10.1016/j.ahj.2009.01.004] [Medline: 19332201]
4. Wall J, White LD, Lee A. Novel ECG changes in acute coronary syndromes. Would improvement in the recognition of "STEMI-equivalents" affect time until reperfusion? Intern Emerg Med 2018 Mar;13(2):243-249. [doi: 10.1007/s11739-016-1595-3] [Medline: 28040835]
5. Tzimas G, Antiochos P, Monney P, et al. Atypical electrocardiographic presentations in need of primary percutaneous coronary intervention. Am J Cardiol 2019 Oct 15;124(8):1305-1314. [doi: 10.1016/j.amjcard.2019.07.027] [Medline: 31455501]
6. Ricci F, Martini C, Scordo DM, et al. ECG patterns of occlusion myocardial infarction: a narrative review. Ann Emerg Med 2025 Apr;85(4):330-340. [doi: 10.1016/j.annemergmed.2024.11.019] [Medline: 39818676]

XSL•FO
RenderX

7. Herman R, Meyers HP, Smith SW, et al. International evaluation of an artificial intelligence-powered electrocardiogram model detecting acute coronary occlusion myocardial infarction. Eur Heart J Digit Health 2023;5(2):123-133. [doi: 10.1093/ehjdh/ztad074] [Medline: 38505483]
8. Al-Zaiti SS, Martin-Gill C, Zègre-Hemsey JK, et al. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. Nat Med 2023 Jul;29(7):1804-1813. [doi: 10.1038/s41591-023-02396-3] [Medline: 37386246]
9. Herman R, Demolder A, Vavrik B, et al. Validation of an automated artificial intelligence system for 12-lead ECG interpretation. J Electrocardiol 2024;82:147-154. [doi: 10.1016/j.jelectrocard.2023.12.009] [Medline: 38154405]
10. Demolder A, Kresnakova V, Hojcka M, et al. High precision ECG digitization using artificial intelligence. J Electrocardiol 2025;90:153900. [doi: 10.1016/j.jelectrocard.2025.153900] [Medline: 40010101]
11. Meyers HP, Bracey A, Lee D, et al. Comparison of the ST-elevation myocardial infarction (STEMI) vs. NSTEMI and occlusion MI (OMI) vs. NOMI paradigms of acute MI. J Emerg Med 2021 Mar;60(3):273-284. [doi: 10.1016/j.jemermed.2020.10.026] [Medline: 33308915]
12. Dixon WC 4th, Wang TY, Dai D, et al. Anatomic distribution of the culprit lesion in patients with non-ST-segment elevation myocardial infarction undergoing percutaneous coronary intervention: findings from the National Cardiovascular Data Registry. J Am Coll Cardiol 2008 Oct 14;52(16):1347-1348. [doi: 10.1016/j.jacc.2008.07.029] [Medline: 18929247]
13. Meyers HP, Smith SW. Prospective, real-world evidence showing the gap between ST elevation myocardial infarction (STEMI) and occlusion MI (OMI). Int J Cardiol 2019 Oct 15;293:48-49. [doi: 10.1016/j.ijcard.2019.07.043] [Medline: 31331673]
14. Hillinger P, Strebel I, Abächerli R, et al. Prospective validation of current quantitative electrocardiographic criteria for ST-elevation myocardial infarction. Int J Cardiol 2019 Oct 1;292:1-12. [doi: 10.1016/j.ijcard.2019.04.041] [Medline: 31056411]

Abbreviations

ACS: acute coronary syndrome
CHUV: Lausanne University Hospital
ECG: electrocardiogram
ED: emergency department
NPV: negative predictive value
NSTEMI: non–ST-elevation myocardial infarction
OMI: occlusion myocardial infarction
PPV: positive predictive value
STEMI: ST-elevation myocardial infarction
TIMI: thrombolysis in myocardial infarction

Edited by A Coristine; submitted 24.Nov.2025; peer-reviewed by A Abdi, M Kokash, N Ainani, S Mohanadas; revised version received 07.Jul.2026; accepted 09.Jul.2026; published 14.Oct.2026.

Please cite as:
Zimmerli A, Dupre M, Liabot Q, Sellah CY, Meier D, Salihu A, Weerts V, Rubimbura V, Roux O, Muller O, Hugli O, Carron PN, Fournier S
Human vs Artificial Intelligence in Detecting Coronary Occlusions in ST-Elevation Myocardial Infarction: Retrospective Study
JMIR Cardio 2026;10:e88399
URL: https://cardio.jmir.org/2026/1/e88399
doi:10.2196/88399

© Aurelia Zimmerli, Marion Dupre, Quentin Liabot, C-Yasmine Sellah, David Meier, Adil Salihu, Victor Weerts, Vladimir Rubimbura, Olivier Roux, Olivier Muller, Olivier Hugli, Pierre-Nicolas Carron, Stephane Fournier. Originally published in JMIR Cardio (https://cardio.jmir.org), 14.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of Rule-Based Natural Language Processing and Diagnosis Code Approaches to Detecting Bleeding in the Electronic Health Record: Algorithm Validation Within a Retrospective Cohort Study

Evan W Draper[1], PharmD; Jan A Anderson[1], PharmD; Surbhi Shah[2], MBBS; Leilia Brown[3], MSN, RN; Margaret R Hanson[4], RN; Kristin D Negley[5], MS, APRN, CNS; Jennifer M Lutz[6], MSN, RN; Andrea E Anderson[1], BSPharm; John G O'Meara[1], PharmD; Nicholas D Jakowenko[1], PharmD; Robert D McBane[7], MD

[1]Department of Pharmacy, Mayo Clinic, 201 1st St SW, Rochester, MN, United States

[2]Division of Hematology and Medical Oncology, Department of Internal Medicine, Mayo Clinic, Phoenix, AZ, United States

[3]Department of Nursing, Mayo Clinic, Jacksonville, FL, United States

[4]Department of Nursing, Mayo Clinic Health System, Mankato, MN, United States

[5]Department of Nursing, Mayo Clinic, Rochester, MN, United States

[6]Department of Nursing, Mayo Clinic, Scottsdale, AZ, United States

[7]Division of Vascular Medicine, Gonda Vascular Center, Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Evan W Draper, PharmD
Department of Pharmacy, Mayo Clinic, 201 1st St SW, Rochester, MN, United States

Abstract

Background: Anticoagulant use is prevalent and increasing. High-quality anticoagulation management balances individual thrombotic and bleeding risks to optimize care. An automated algorithm to accurately extract major bleeding events from the electronic health record (EHR) of a large health care system would be extremely useful for tracking outcomes of anticoagulated patients.

Objective: The study aims to assess the accuracy of a natural language processing (NLP) approach, compared with and in conjunction with a diagnosis code (Dx Code)–based strategy, for detecting major bleeding from the EHR.

Methods: Patients with at least 1 prescription for apixaban, dabigatran, edoxaban, rivaroxaban, or warfarin between January 1, 2019, and December 31, 2023, were analyzed for major bleeding events. Major bleeding was defined similarly to the International Society on Thrombosis and Haemostasis definition. For the NLP algorithm, clinical notes and radiology reports were screened for mentions of bleeding. MedTagger, an open-source, rule-based NLP tool, was then used on the notes. The NLP results were transformed into a feature set, which was used to create a support vector machine model. Three NLP models were created for central nervous system bleeding, critical site bleeding, and bleeding with a hemoglobin drop or transfusions (other major bleeding). Training and test datasets were created using an active learning approach. The model was trained using 80% of the dataset, while the remaining 20% was used for testing. After model finalization, an independent random sample was used to validate the model. Dx Codes were also used to detect major bleeding, similar to previously published approaches. Five unique algorithms were compared: (1) NLP, (2) Dx Code in the primary position, (3) Dx Code in any position, (4) NLP∩Dx Code in any position, and (5) NLP∪Dx Code in any position. All 5 algorithms were compared with manual chart review.

Results: The final analysis identified 111,338 patients prescribed outpatient anticoagulation during the study. Manual adjudication was performed for 1189 of the 111,338 (1%) patients. Most of the 5 approaches performed well on the test data, with F_1-scores ranging from 0.545 to 0.959, except for Dx Code–Primary, with an F_1-score of 0.353 for critical site bleeding. The F_1-scores for the validation data ranged from 0.509 to 0.679. The NLP∩Dx Code in any position approach yielded excellent precision of 0.864 (95% CI 0.667 - 0.953) but only 0.559 (95% CI 0.395 - 0.711) sensitivity. The NLP∪Dx Code in any position approach yielded good sensitivity of 0.824 (95% CI 0.665 - 0.917) but only 0.368 (95% CI 0.269 - 0.481) precision.

Conclusions: Electronic tools can screen the EHRs to identify major bleeding outcomes in large patient cohorts within a multisite health care network with high precision but only moderate sensitivity. Additional work is needed to enhance sensitivity while maintaining high precision.

(JMIR Cardio 2026;10:e87424) doi:10.2196/87424

XSL•FO
RenderX

KEYWORDS

anticoagulants; hemorrhage; machine learning algorithms; natural language processing; population surveillance

Introduction

A growing number of patients take antithrombotic therapy, including more than 8 million on anticoagulants in the United States and more than 25 million on antiplatelet agents [1,2]. Overtime, this number is only anticipated to grow with the aging population. These medications carry a risk of major hemorrhage, which adversely impacts quality of life and increases mortality risk. There are an estimated 250,000 anticoagulant-related emergency department (ED) visits annually for bleeding outcomes and for every 100 patients dispensed an anticoagulant, there are nearly 9 ED visits [3]. High-quality anticoagulation delivery requires individualized patient assessment that balances thrombotic and bleeding risks to optimize management. This decision-making process is nuanced and dynamic, with overlap between bleeding and thrombotic risk factors as patients age.

High-quality delivery of anticoagulation for large health care networks involving tens of thousands of patients can be challenging. When considering ideal oversight of large anticoagulation networks, tools capable of monitoring hard outcomes are essential for quality improvement, and central to this process is an accurate and near-real–time assessment of bleeding events.

Given the widespread use of antithrombotic agents, manual chart abstraction for bleeding outcomes ascertainment is neither timely, cost-effective, nor practical. Methodologies using diagnosis code (Dx Code)–based major bleeding detection are widely used [4-8] but may lack sensitivity, specificity, and overall accuracy [9-11]. Electronic tools capable of automated and near-real–time identification of bleeding events could facilitate quality improvement and the refinement of safer anticoagulation delivery.

Established in 2019, the Mayo Clinic Anticoagulation and Antiplatelet Stewardship Quality Committee has been charged with oversight of antithrombotic delivery across the multisite institution. Since 2019, this has included more than 115,000 patients receiving a direct oral anticoagulant and more than 40,000 patients treated with warfarin. To facilitate the management of this large patient population, tools capable of providing near-real–time detection of bleeding and thrombosis are required. The objective of this study was to assess the accuracy of a natural language processing (NLP) approach compared with and in conjunction with a Dx Code–based strategy for detecting major bleeding.

Methods

Patient Population

An algorithm validation within a retrospective cohort study design was used to evaluate the accuracy of NLP, Dx Codes, and NLP plus Dx Codes to identify major bleeding in the electronic health record (EHR). Patients with at least one outpatient prescription for apixaban, dabigatran, edoxaban, rivaroxaban, or warfarin (January 1, 2019, to December 31, 2023) were included in the analysis. Data were collected from a centralized EHR that contains detailed information about all inpatient hospitalizations, outpatient visits, radiology images and reports, laboratory results, pathology results, and patient correspondence. The indication for anticoagulation was estimated based on Dx Codes in the EHR. Demographic information, indication, and prescribed medication were provided for all manually reviewed patients. Any missing values were categorized as unknown.

Bleeding Definition

Major bleeding was defined similar to the ISTH (International Society on Thrombosis and Haemostasis) and included symptomatic bleeding into a critical area or organ, and bleeding causing a hemoglobin decrease of at least 2 g/dL or transfusions of at least 2 units of packed red blood cells (PRBC) [12]. Symptomatic bleeding into a critical area or organ was divided into 2 categories: central nervous system (CNS) bleeding and critical site bleeding. CNS bleeds were defined as either intracranial or intraspinal bleeding. Critical site bleeding included intraocular, retroperitoneal, pericardial, hemothorax [13,14], and intra-articular bleeding. Bleeding with a hemoglobin drop or transfusions was referred to as “other major bleeding.” Fatal bleeding is included in the ISTH bleeding definition. However, a Dx Code specific to fatal bleeding does not exist, making it impossible to evaluate specifically for fatal bleeding without assuming the cause of death or manually reviewing the chart. Fatal bleeding events are likely to be found by the NLP approach and Dx Codes, but neither approach will specifically label the event as a fatal bleed. Mortality occurring within the same month as a bleeding event was collected and is described.

The ISTH definition of symptomatic bleeding with a 2 g/dL or greater change in hemoglobin or transfusion of 2 or more units of PRBC (“other major bleeding”) does not specify the timing for comparing hemoglobin results or for PRBC transfusions. For the purposes of this study, a qualifying drop in hemoglobin was defined by 2 methods. The first method compared a hemoglobin result with any other hemoglobin result within the preceding 30 days. The second method compared a hemoglobin result to the immediately prior hemoglobin result. If either method was true, then the hemoglobin change met the definition. Using both approaches captures a slow, occult bleed with months between hemoglobin results as well as a more acute drop in hemoglobin values. For the transfusion requirement definition, 2 or more units of PRBC had to be administered within a 3-day time increment. If there was a 4-day separation between transfusions, that would not meet the definition. Other studies have proposed a 48-hour limit on the hemoglobin change and did not specify a time frame for the transfusions [15]. The temporal thresholds for transfusions and hemoglobin change for this study were determined through active learning and by examining bleeding events within the EHR.

XSL•FO
RenderX

Event Identification and Adjudication

To compare the accuracy of NLP and Dx Codes to detect bleeding outcomes, manual chart review for event validation was performed. Positive bleeding events for training, testing, and validation were manually and independently reviewed in the EHR by at least 2 study team members (nurses, pharmacists, or physicians) to create the gold standard against which the NLP and Dx Code approaches were compared. Months in which no mentions of bleeding terms were noted throughout the EHR were assumed to be months without a bleeding event, and no EHR manual review was completed. Months in which bleeding terms were found or Dx Codes were present were manually reviewed in the EHR by 1 study team member. Obvious nonevents, such as "no intracranial hemorrhage present," were categorized as nonevents by only 1 study team member. Ambiguous situations such as "differential diagnosis includes hemarthrosis, septic arthritis…" were reviewed by at least 2 study team members when the initial review could not confidently rule out a bleeding event. The date and type of the bleed (CNS, critical site, or other major bleed) were recorded, and each month adjudicated was classified as a bleeding event or not a bleeding event. Adjudication by month was chosen after an initial 3-day rolling window proved performance-intensive. Cohen κ was calculated by evaluating interteam member rating reliability on the validation sample. Specific sentences or notes were not annotated for NLP purposes. A bleeding event could occur in the outpatient or inpatient setting. Event disagreements were resolved through group discussion. Only the incident bleeding event was included for patients with multiple events. While prescribed anticoagulation was an inclusion criterion for the analysis, temporary anticoagulant interruption prior to the bleeding event did not exclude the patient. For event adjudication, surgical bleeding leading to a drop in hemoglobin or transfusion was excluded. The NLP algorithm does not have the functionality to exclude surgical bleeding, and future work would be required to align the algorithm with manual review. Additional details on event adjudication are available in the Section S1 of Multimedia Appendix 1.

NLP

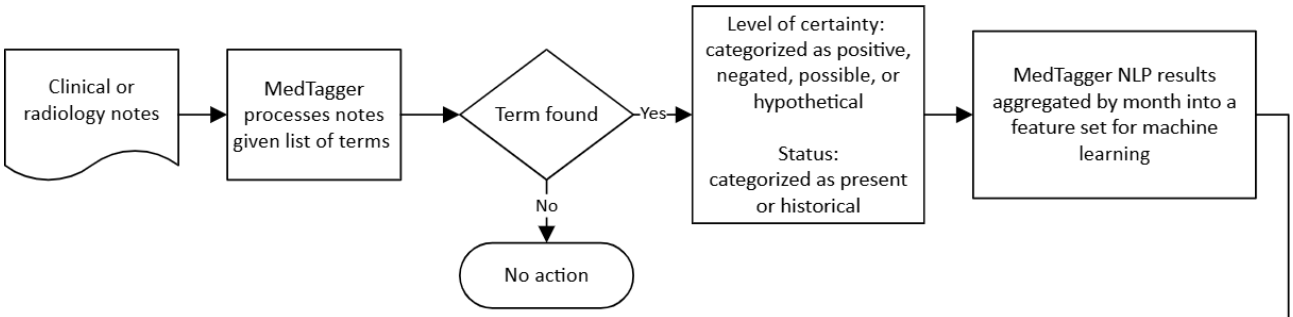
MedTagger is an open-source, rule-based NLP program from Open Health Natural Language Processing (OHNLP) Consortium that detects specific terms and then uses more than 700 rules to determine the context of each term [16]. MedTagger classifies the context using the following elements: experiencer: patient or other individual; status: present or historical; and level of certainty: positive, negative, possible, or hypothetical. The text was not preprocessed before using MedTagger. A MedTagger version based on the OHNLP Toolkit (v1.0.82) was used.

All patient notes were searched for CNS bleeding and critical site bleeding using regular expressions. The incident date and 3 days following a 2 g/dL or greater drop in hemoglobin or 2 or more units of PRBC were also searched using regular expressions. Search terms were specific to each bleeding type. Additional information on the search terms is provided in Section S2 of Multimedia Appendix 1. Notes with bleeding mentions found by regular expressions were further processed with MedTagger NLP. Exclusionary terms were added to the algorithm for NLP refinement. Additional details on MedTagger search terms and excluded note sections are available in Section S3 of Multimedia Appendix 1.

A feature set for each algorithm was created, and machine learning using a support vector machine model was created to identify bleeds. Support vector machine was selected as the model after evaluating logistic regression, random forest, and gradient-boosted machine models. Adjudicated data were divided into training (80%) and testing sets (20%), stratified by patient to ensure no data-leakage. Months without a bleeding event were combined with months with a CNS bleed and used to create the CNS bleed model. The same months without a bleeding event were used to create the critical site bleed and other major bleed models. The feature set aggregated NLP results by month, and NLP results from clinical notes and radiology notes were handled as separate features. Machine learning parameters were chosen using grid search and cross-validation on the training set. Additional information on model training is provided in Section S4 of Multimedia Appendix 1. A visual representation of the approach is provided in Figure 1.

XSL•FO
RenderX

Figure 1. MedTagger process. MedTagger is an open-source, rule-based natural language processing (NLP) tool that categorizes medical terms with regard to certainty (ie, positive, negated, or hypothetical) and status (ie, present or history of). This work used MedTagger to categorize mentions of bleeding found in clinical and radiology notes in an attempt to find major bleeding events among patients on anticoagulation. The MedTagger results are aggregated to create a feature set, which is subsequently used in machine learning. MRN: medical record number; *each combination of NLP results creates an additional feature; **one row per MRN per month.

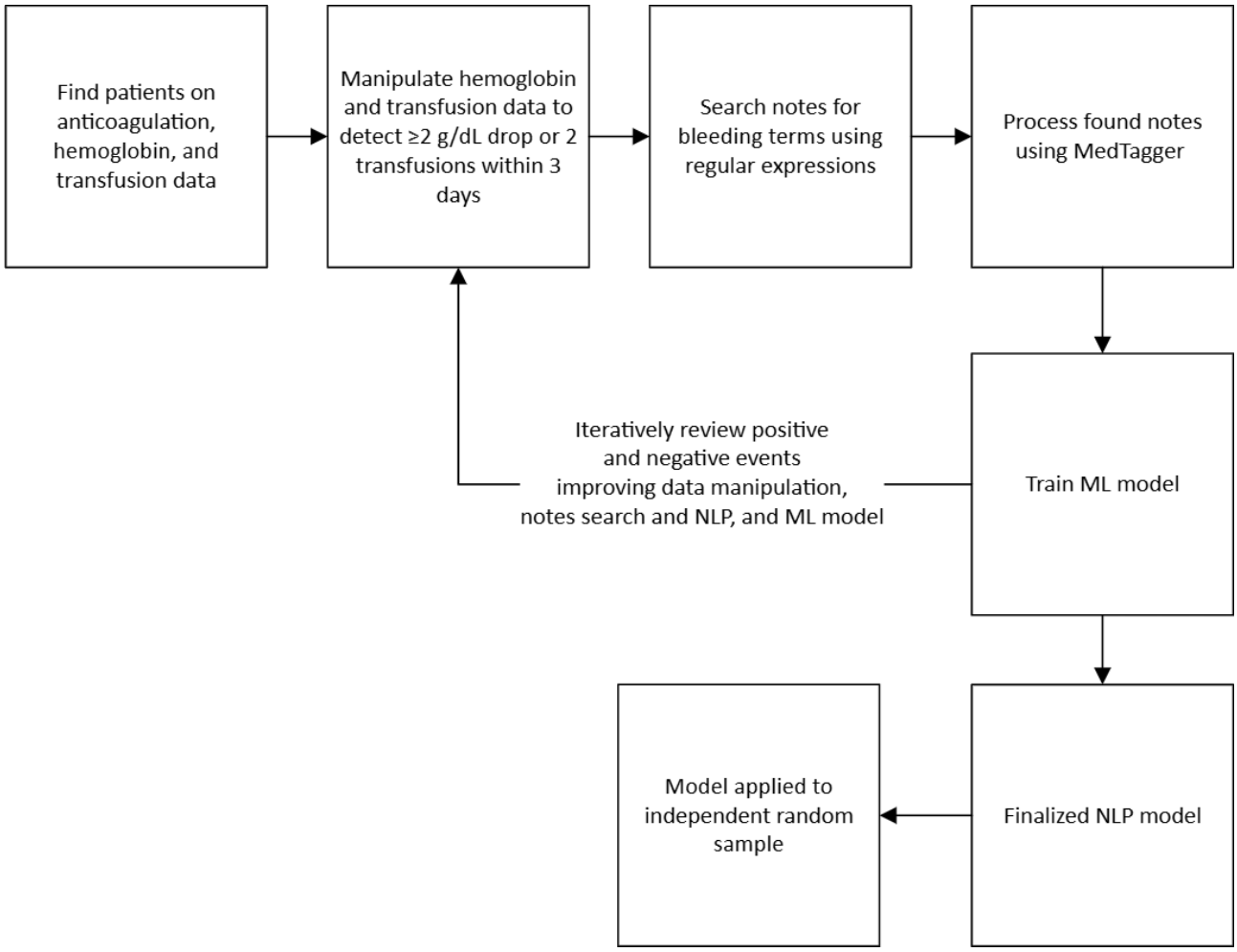


MRN	Month start date	Positive present radiology	Negated present radiology	Positive history of radiology	...*	Positive present clinical
1234	5/1/2021	1	0	0	...	2
1234	4/1/2020	2	0	0	...	1
1234	4/1/2022	1	0	0	...	1
1234	4/1/2021	1	0	0	...	0
1234	7/1/2021	17	0	0	...	0
...**	...	...	...	...	...	...

The NLP approach was refined iteratively by reviewing high and low probability cases similar to active learning (Figure 2) [17]. This approach was used to select cases for manual review and adjudication during model development. This approach does not suffer as much from class imbalance compared to random sampling. After model development and finalization, an independent random sample of patients was manually adjudicated and run through the model.

XSL•FO
RenderX

Figure 2. Natural language processing (NLP) model development. This retrospective cohort study of Mayo Clinic patients prescribed oral anticoagulation from January 1, 2019, to December 31, 2023, assessed the feasibility and accuracy of an NLP and machine learning (ML) approach to detect major bleeding from the electronic health record (EHR). The model development process iteratively refined the search terms, the method of combining discrete hemoglobin and transfusion data, and the feature creation process. The finalized NLP model was applied to an independent, random sample and was compared with 2 different diagnosis code approaches.



Dx Code

Bleeding events were also identified using Dx Codes [8]. Dx Codes associated with ED visits or hospital admissions were included. The primary Dx Code ("Dx Code–Primary") or any Dx Code associated with the encounter ("Dx Code–Any Position") were evaluated. Specific *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) codes used are provided in Section S5 of Multimedia Appendix 1. Data were aggregated by month based on the admission date.

NLP Combined With Dx Code

The Dx Code–Any Position approach and the NLP approach were combined in 2 ways: "NLP∩Dx Code Any" requires both the NLP and Dx Code–Any Position results to be positive, while "NLP∪Dx Code Any" allows either the NLP or the Dx Code–Any Position approach to be positive.

Algorithm and Comparison Summary

Five approaches for three different bleeding types were created: (1) NLP, (2) Dx Code–Primary, (3) Dx Code–Any, (4) NLP∩Dx Code Any, and (5) NLP∪Dx Code Any. For model test data, the 5 approaches were evaluated for each of the 3 different bleeding types (CNS bleeding, critical site bleeding, and other major bleeding). For model validation on a random sample of data, the 3 bleeding types were combined due to the low volume of events.

Statistical Analysis

Accuracy measures of sensitivity, specificity, precision, and F_1-score, with 95% CIs, were calculated for each approach compared against the gold standard of a manually adjudicated review. The methodology to calculate the 95% CIs was SciPy's implementation of the Wilson without continuity correction for sensitivity, specificity, and precision [18]. Given no clear best approach for calculating a CI for F_1-score [19], a bootstrap methodology using 5000 samples was used to calculate the 95% CI for F_1-score. The F_1-score is a measurement that evaluates both sensitivity and precision in a single metric. Cohen κ was calculated by evaluating interteam member rating reliability and used the bootstrap methodology. The above analyses were performed for each type of bleeding as well as for a combined "all bleeding" analysis. All analyses were performed using Python (version 3.10.18).

XSL•FO
RenderX

Ethical Considerations

The project was deemed exempt by the Mayo Clinic Institutional Review Board in the context of a quality improvement project. Patient information was stored in a secure environment within the Mayo Clinic Cloud. Study team members were provided with access to the minimum amount of identifiable information necessary to verify the bleeding event during event adjudication.

Results

The final analysis identified 111,338 total patients prescribed outpatient anticoagulation over the study time frame. Clinical notes and radiology reports for the included patients were searched for bleeding terms, identifying 264,551 clinical notes and radiology reports for CNS bleeds, 73,973 for critical site bleeds, and 581,120 for other major bleeding. The total number of unique notes with bleeding terms found is 773,779. These notes and reports were processed using MedTagger (Figure 3). A cohort of 604 patients was manually adjudicated during active learning and was used for model training and testing. Each patient and month reviewed contributed to model development. A total of 1001 patient-months were manually reviewed, with 643 months without a bleeding event, 183 months with CNS bleeding, 68 months with critical site bleeding, and 107 months with other major bleeding. An additional random sample of 585 patients was used to validate the final model. The validation sample yielded a total of 30,757 patient-months, with 30,723 months without a bleeding event, 7 months with a CNS bleed, 7 months with a critical site bleed, and 20 months with other major bleed. Cohen κ between the reviewers for the random sample was 0.66 (95% CI 0.38-0.88). Demographic information, indications, and medications prescribed for all patients manually reviewed are provided in Table 1. The patients had a mean age of 70 (SD 14.4) years, and 57% were male patients. The majority of patients were White individuals. The 2 most common indications for anticoagulant therapy were atrial fibrillation and thromboembolism. The 2 most common anticoagulants used were apixaban and warfarin.

XSL•FO
RenderX

Figure 3. Pipeline and adjudicated reviews flow diagram. A natural language processing (NLP) and machine learning–based approach to detect major bleeding events from the electronic health record (EHR) was developed and evaluated. The approach processed 111,338 patients and 773,779 unique notes. Patients on oral anticoagulation were first identified. Notes mentioning bleeding were found and processed by MedTagger. Manual reviews of bleeding events were performed to evaluate the accuracy of the approach. CNS: central nervous system.

XSL•FO
RenderX

Table . Baseline demographics and clinical characteristics.

Characteristic	Patients (N=1189)
Age (y), mean (SD)	70.0 (14.4)
Sex: female, n (%)	514 (43.2)
Race, n (%)	
White	1108 (93.2)
Black	29 (2.4)
Asian	20 (1.7)
American Indian/Alaskan Native	5 (0.4)
Other/unknown	27 (2.3)
Ethnicity, n (%)	
Not Hispanic/Latino	1142 (96.0)
Hispanic/Latino	25 (2.1)
Not disclosed/unknown	22 (1.9)
Anticoagulant indication, n (%)	
Atrial fibrillation	638 (53.7)
Thromboembolism	399 (33.6)
Other/unknown	111 (9.3)
Heart valve prosthesis or repair	23 (1.9)
Peripheral artery disease	18 (1.5)
Anticoagulant used, n (%)	
Apixaban	630 (53.0)
Warfarin	342 (28.8)
Rivaroxaban	208 (17.5)
Dabigatran	8 (0.7)
Edoxaban	1 (0.1)

Performance metrics and 95% CIs for the testing data and individual bleeding types (Table 2) compare the performance of the 5 different approaches. CNS bleeding algorithms performed the best with F_1-scores of 0.946 (95% CI 0.885 - 0.989) for NLP, 0.959 (95% CI 0.904 - 1.000) for Dx Code–Primary, and 0.947 (95% CI 0.886 - 0.989) for Dx Code–Any, 0.959 (95% CI 0.906 - 1.000) for NLP∩Dx Code Any, and 0.935 (95% CI 0.868 - 0.986) for NLP∪Dx Code Any.

XSL•FO
RenderX

Table . Performance metrics for individual bleeding types for test data.

Bleeding classification method	Sensitivity (95% CI)	Specificity (95% CI)	Precision (95% CI)	F_1-score (95% CI)
CNS[a] bleeds				
NLP[b]	0.946 (0.823 - 0.985)	0.982 (0.938 - 0.995)	0.946 (0.823 - 0.985)	0.946 (0.885 - 0.989)
Dx Code[c]–Primary	0.946 (0.823 - 0.985)	0.991 (0.952 - 0.998)	0.972 (0.858 - 0.995)	0.959 (0.904 - 1.000)
Dx Code–Any	0.973 (0.862 - 0.995)	0.974 (0.925 - 0.991)	0.923 (0.797 - 0.973)	0.947 (0.886 - 0.989)
NLP∩Dx Code Any[d]	0.946 (0.823 - 0.985)	0.991 (0.952 - 0.998)	0.972 (0.858 - 0.995)	0.959 (0.906 - 1.000)
NLP∪Dx Code Any[e]	0.973 (0.862 - 0.995)	0.965 (0.913 - 0.986)	0.900 (0.769 - 0.960)	0.935 (0.868 - 0.986)
Critical site bleeds				
NLP	0.643 (0.388 - 0.837)	0.983 (0.942 - 0.995)	0.818 (0.523 - 0.949)	0.720 (0.471 - 0.897)
Dx Code–Primary	0.214 (0.076 - 0.476)	1.000 (0.969 - 1.000)	1.000 (0.439 - 1.000)	0.353 (0.000 - 0.632)
Dx Code–Any	0.429 (0.214 - 0.674)	0.983 (0.942 - 0.995)	0.750 (0.409 - 0.929)	0.545 (0.235 - 0.769)
NLP∩Dx Code Any	0.429 (0.214 - 0.674)	1.000 (0.969 - 1.000)	1.000 (0.610 - 1.000)	0.600 (0.286 - 0.828)
NLP∪Dx Code Any	0.643 (0.388 - 0.837)	0.967 (0.918 - 0.987)	0.692 (0.424 - 0.873)	0.667 (0.421 - 0.857)
Other major bleeds				
NLP	0.818 (0.615 - 0.927)	0.958 (0.906 - 0.982)	0.783 (0.581 - 0.903)	0.800 (0.650 - 0.913)
Dx Code–Primary	0.545 (0.347 - 0.731)	0.975 (0.929 - 0.991)	0.800 (0.548 - 0.930)	0.649 (0.437 - 0.809)
Dx Code–Any	0.636 (0.43 - 0.803)	0.942 (0.884 - 0.971)	0.667 (0.454 - 0.828)	0.651 (0.452 - 0.800)
NLP∩Dx Code Any	0.591 (0.387 - 0.767)	0.983 (0.941 - 0.995)	0.867 (0.621 - 0.963)	0.703 (0.500 - 0.851)
NLP∪Dx Code Any	0.864 (0.667 - 0.953)	0.917 (0.853 - 0.954)	0.655 (0.473 - 0.801)	0.745 (0.595 - 0.864)

[a]CNS: central nervous system.

[b]NLP: natural language processing.

[c]Dx Code: diagnosis code.

[d]NLP∩Dx Code Any is the intersection (∩) of NLP and Dx Code Any, where both approaches must be positive to qualify as a bleed.

[e]NLP∪Dx Code Any is the union (∪) of NLP and Dx Codes where either approach can be positive to qualify as a bleed.

Based on F_1-scores, the NLP approach performed the best for critical site bleeds, with 0.720 (95% CI 0.471 - 0.897), and for other major bleeds, with a score of 0.800 (95% CI 0.650 - 0.913).

Performance metrics for the validation data and combined bleeds (Table 3) demonstrate decreased performance compared with the test data. F_1-scores ranged from 0.509 (95% CI 0.388-0.618) to 0.679 (95% CI 0.524-0.812) for validation compared with up to 0.959 for the test data. The best precision was for the NLP∩Dx Code Any with 0.864 (95% CI 0.667 - 0.953). However, this approach only yielded 0.559 (95% CI 0.395 - 0.711) sensitivity. Model features are available in Section S6 of Multimedia Appendix 1.

XSL•FO
RenderX

Table . Performance metrics for the random validation sample.

Bleeding identification method	Sensitivity (95% CI)	Specificity (95% CI)	Precision (95% CI)	F_1-score (95% CI)
NLP[a]	0.676 (0.508 - 0.809)	0.999 (0.999 - 1.000)	0.511 (0.370 - 0.650)	0.582 (0.441 - 0.703)
Dx Code[b]–Primary	0.618 (0.450 - 0.761)	1.000 (0.999 - 1.000)	0.618 (0.450 - 0.761)	0.618 (0.462 - 0.738)
Dx Code–Any	0.706 (0.538 - 0.832)	0.999 (0.999 - 0.999)	0.453 (0.327 - 0.585)	0.552 (0.415 - 0.667)
NLP∩Dx Code Any[c]	0.559 (0.395 - 0.711)	1.000 (1.000 - 1.000)[d]	0.864 (0.667 - 0.953)	0.679 (0.524 - 0.812)
NLP∪Dx Code Any[e]	0.824 (0.665 - 0.917)	0.998 (0.998 - 0.999)	0.368 (0.269 - 0.481)	0.509 (0.388 - 0.618)

[a]NLP: natural language processing.

[b]Dx Code: diagnosis code.

[c]NLP∩Dx Code Any is the intersection (∩) of NLP and Dx Code Any, where both approaches must be positive to qualify as a bleed.

[d]The specificity, rounded to 6 decimal places, is 0.999902 (0.999713-0.999967).

[e]NLP∪Dx Code Any is the union (∪) of NLP and Dx Code Any, where either approach can be positive to qualify as a bleed.

Mortality in the training and test data occurred within the same month as the bleed outcome for 22 out of 183 (12%) CNS bleeds, 2 out of 68 (3%) critical site bleeds, and 5 out of 107 (5%) other major bleeds. Mortality in the validation data occurred within the same month, 2 out of 7 (29%) of CNS bleeds, 0 out of 7 (0%) critical site bleeds, and 2 out of 20 (10%) other major bleeds. While causality assignment for fatal bleeding can be challenging, the temporal association between the bleeding event and death suggests a potential contribution to this outcome.

Discussion

Principal Findings

The principal finding of the current study is that electronic tools can screen electronic medical records for major bleeding outcomes from large patient cohorts within a multisite health care network with high precision and moderate sensitivity. In validation, the NLP∩Dx Code Any approach may have the most clinical utility with the highest precision of 0.864 (95% CI 0.667 - 0.953) and moderate sensitivity of 0.559 (95% CI 0.395 - 0.711). Given the high volume of patients on anticoagulant medications, an electronic tool that maximizes precision might provide the most useful measure for groups focused on quality improvement and process oversight.

The clinical utility of these findings is notable for several reasons. One could envision the use of this information to improve patient outcomes. First, appropriate use criteria could be assessed to ensure anticoagulants are appropriately prescribed. For example, patients with severe renal impairment, liver impairment, or thrombocytopenia may be at particular risk for bleeding outcomes. When tracked electronically, "our practice alerts" could then be delivered to providers. Second, duplicate antithrombotic delivery could be scrutinized and corrected. This could include discontinuation of antiplatelets where appropriate. Third, dose adjustments of direct anticoagulants could be reviewed such as the "ABC" criteria for apixaban or dose reductions for secondary venous thromboembolism prevention after 6 months. Fourth, this tool could be combined with electronic measures of International Normalized Ratio (INR) stability for those patients receiving warfarin therapy. While engaging the practice after a major bleed or intracranial bleed has occurred is not ideal, the use of this tool could have a major impact on practice change to prevent the next bleeding event. Fifth, these tools can provide a real-world, in-house assessment of new antithrombotic strategies. As the factor XI inhibitors are Food and Drug Administration (FDA)–approved and incorporated into clinical practice, adverse outcomes, compared with traditional anticoagulants, can be assessed in real-time. Clearly, more work is needed to refine this and other tools to maximize utility. These types of tools may then enable incorporation into clinical practice for quality oversight.

Comparison to Prior Work

Multiple recent studies have evaluated the ability to detect and categorize bleeding events using structured [15,20], unstructured [21], or a combination of structured and unstructured data [22,23]. Coumau et al [23] evaluated various approaches for detecting major bleeding as well as clinically relevant nonmajor bleeding. Their top approach was a combination of structured and unstructured data and yielded a sensitivity of 0.84, a precision of 0.51, and an F_1-score of 0.64. Gaspar et al [21] performed document-level categorization of discharge summaries of older adults and compared an NLP approach with a rule-based, Dx Code approach. The NLP approach outperformed the rule-based, Dx Code approach in the ability to categorize a document as "no bleeding," "clinically significant bleeding," "severe bleeding," and "history of bleeding" with an F_1-score of 0.80. An et al [20] applied a previously published algorithm [15] using structured data to detect major bleeding and clinically relevant nonmajor bleeding. They report a sensitivity of 0.91 and specificity of 0.92 for detecting major bleeding and a sensitivity of 0.52 and specificity of 0.92 for detecting clinically relevant nonmajor bleeding. They used a conditional sampling approach for the validation.

Older studies have used similar approaches to scrutinize clinical outcomes among large cohorts of anticoagulated patients. Cunningham et al [9] used Dx Codes in addition to transfusion data to identify bleeding related to anticoagulation and achieved an 89% positive predictive value. Negative predictive value was not reported. Arnason et al [10] reviewed patient charts with a

XSL•FO
RenderX

"bleeding" Dx Code and determined if a major bleed occurred. Accuracy measures with 95% CIs were reported. Definite bleeding had a sensitivity of 93% (95% CI 90% - 96%) and specificity of 88% (95% CI 83% - 91%). Major bleeding had a sensitivity of 94% (95% CI 91% - 96%) and specificity of 83% (95% CI 78% - 87%). Tirschwell and Longstreth [11] used Dx Codes to detect hemorrhagic stroke in administrative claims data, including primary diagnosis only, first 2 diagnoses, and all diagnoses. The best algorithm differed by stroke type, and for intracranial hemorrhage and subarachnoid hemorrhage, using only the primary diagnosis maximized agreement. Intracranial hemorrhage had sensitivity of 82% (95% CI 66% - 92%), specificity of 93% (95% CI 86% - 97%), and κ 0.74 (0.58 - 0.90), while subarachnoid hemorrhage had a sensitivity of 98% (95% CI 90% - 100%), specificity of 92% (95% CI 84% - 96%), and κ 0.87 (95% CI 0.71 - 1.00). While Dx Code approaches can be automated, they may not have sufficient accuracy metrics to be used in practice or within a learning health system. For example, a method with a sensitivity of 98% and a specificity of 92% applied to a population of 100,000 and assuming a 1% bleeding rate would result in 9780 false positives and 20 false negatives and a precision of 11%.

Others have assessed strategies that do not use Dx Codes for bleeding detection. Lee et al [24] studied bleeding events related to clopidogrel usage. Ten different approaches, 3 rule-based, 3 score-based, and 4 machine learning–based approaches were analyzed. While the highest precision was 97%, the sensitivity was low at 29%. A machine learning approach performed best, with a precision of 89%, a sensitivity of 97% (sensitivity in this study may not be accurate due to nonrandom sampling), and an F_1-score of 92%. Taggart et al [25] searched for "clinically relevant bleeding" in the general adult intensive care unit population using the MIMIC-III (Medical Information Mart for Intensive Care) database. Using NLP results, they created rule-based and machine learning approaches for detecting bleeding. "Clinically relevant bleeding" was defined as "as symptomatic bleeding that required treatment." The rule-based approach had a sensitivity of 91%, specificity of 85%, and precision of 63%. The machine learning–based approach had a sensitivity of 94%, specificity of 72%, and precision of 49%. Shung et al [26] assessed gastrointestinal bleeding in the ED. An early large language model (LLM) was the best approach (sensitivity of 93%, specificity of 63%, and precision of 84%) compared with either an NLP machine learning model and a discrete data model. More recent studies have focused on bleeding mention detection at the sentence or note level [27,28]. Further studies are needed to determine how sentence or note level detection would translate into patient level detection of bleeding events.

LLMs for information extraction present a potential future step beyond rule-based NLP that may have a superior ability to distinguish between bleeding events and no bleeding. Limitations of LLM approaches include potentially high cost, processing time and power requirements, and data privacy concerns. While recent advances suggest that LLMs and deep learning language models may be superior to rule-based NLP [29-31], 1 recent study found that rule-based NLP remained superior [32]. One potential future direction of this work is to compare and/or combine the NLP∩Dx Code Any approach with an LLM approach.

Study strengths include the use of a rule-based NLP system to detect major bleeding for over 100,000 patients prescribed outpatient anticoagulation. This proof of concept demonstrates that large-scale detection of bleeding with NLP is feasible. Train, test, and validation cases were adjudicated by multiple reviewers to ensure accuracy. Two commonly used approaches were compared with a novel NLP approach that combines structured data of hemoglobin and transfusions to approximate the ISTH major bleed definition and scrutinizes both clinical and radiology notes for bleeding detection. An open-source NLP program that can be applied to large numbers of patients and notes was used at minimal expense.

Limitations

The study has several limitations. First, only the incident bleeding event per patient was included. Yet, it is well known that patients can experience more than one bleeding event. Indeed, those patients with recurrent bleeding may be fundamentally different compared to those with an isolated event. By restricting the outcome to the initial bleed without further refinement, important clinical variables may be missed. Second, while considerable time was spent to thoroughly validate 1189 patients, this represents only 1% of the total population at risk. Whether this was an adequately representative sample is not known. Third, patients who suffered a major bleed were not necessarily on anticoagulation at the time. There may be fundamental differences between anticoagulation-related major bleeding and bleeding events occurring without anticoagulation. Fourth, in the manual adjudication, patients experiencing a bleeding complication related to surgery were excluded. These represent a unique and important cohort; however, uncontrollable variables related to surgical technique, surgical site, and extent are difficult to control for and unique to surgical populations. Surgical bleeding is a topic unto itself and not easily comprehensible when combined with nonsurgical bleeding. Fifth, the approach was not externally validated and relies on infrastructure that may be unique to Mayo Clinic. Sixth, the Cohen κ metric on the validation sample was 0.66 (95% CI 0.38 - 0.88), which is lower than desired. Adjudicating bleeding events is challenging, and discrepancies between reviewers were resolved by physician review or group consensus.

Conclusions

A rule-based NLP pipeline processing tool was developed and validated for analysis of more than 100,000 patients over a period of 5 years. These kinds of tools are essential for developing quality improvement programs with the structured delivery of high-quality anticoagulation for large patient populations. While the "NLP∩Dx Code" in any position approach yielded excellent precision, the sensitivity of this approach needs to be improved. Future directions will combine this rule-based and Dx Code approach with new LLM-based approaches to information extraction.

XSL•FO
RenderX

Acknowledgments

Che Ngufor and Shrinath Patel reviewed machine learning scripts and provided technical guidance. David Hodge reviewed the statistical plan.

Generative AI Use

Generative AI was not used in the drafting of this manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the presence of protected health information (PHI) in clinical notes and radiology reports. Summary counts of natural language processing findings and/or code are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: EWD, SS, RDM
Data curation: EWD (lead), JAA, SS, LB, MRH, KDN, JML, AEA, JGO, NDJ, RDM
Formal analysis: EWD
Investigation: EWD
Supervision: SS, RDM
Writing – original draft: EWD
Writing – review & editing: EWD, JAA, SS, LB, MRH, KDN, JML, AEA, JGO, NDJ, RDM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information regarding event adjudication, note search terms, MedTagger search terms, model training, diagnosis codes, and feature lists.
[DOCX File, 37 KB - cardio_v10i1e87424_app1.docx]

References

1. Yates NY, Hale SA, Clark NP. The impact of clinical pharmacy services on direct oral anticoagulant medication selection and dosing in the ambulatory care setting. J Pharm Pract 2024 Jun;37(3):671-676. [doi: 10.1177/08971900231166555] [Medline: 36989436]
2. Gupta M, Gulati S, Nasir K, Sarraju A. Aspirin use prevalence for cardiovascular disease prevention among U.S. adults from 2012 to 2021. Ann Intern Med 2024 Aug;177(8):1139-1141. [doi: 10.7326/M24-0427] [Medline: 38914002]
3. Geller AI, Shehab N, Lovegrove MC, Weidle NJ, Budnitz DS. Bleeding related to oral anticoagulants: trends in US emergency department visits, 2016-2020. Thromb Res 2023 May;225:110-115. [doi: 10.1016/j.thromres.2023.03.010] [Medline: 37062120]
4. Siontis KC, Zhang X, Eckard A, et al. Outcomes associated with apixaban use in patients with end-stage kidney disease and atrial fibrillation in the United States. Circulation 2018 Oct 9;138(15):1519-1529. [doi: 10.1161/CIRCULATIONAHA.118.035418] [Medline: 29954737]
5. Reynolds MR, Gunnarsson CL, Hunter TD, et al. Health outcomes with catheter ablation or antiarrhythmic drug therapy in atrial fibrillation: results of a propensity-matched analysis. Circ Cardiovasc Qual Outcomes 2012 Mar 1;5(2):171-181. [doi: 10.1161/CIRCOUTCOMES.111.963108] [Medline: 22373904]
6. Jun M, James MT, Manns BJ, et al. The association between kidney function and major bleeding in older adults with atrial fibrillation starting warfarin treatment: population based observational study. BMJ 2015 Feb 3;350:h246. [doi: 10.1136/bmj.h246] [Medline: 25647223]
7. White RH, Beyth RJ, Zhou H, Romano PS. Major bleeding after hospitalization for deep-venous thrombosis. Am J Med 1999 Nov;107(5):414-424. [doi: 10.1016/s0002-9343(99)00267-3] [Medline: 10569295]
8. Ngufor C, Yao X, Inselman JW, et al. Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method. Am Heart J 2023 Jun;260:124-140. [doi: 10.1016/j.ahj.2023.02.015] [Medline: 36893934]

XSL•FO
RenderX

9. Cunningham A, Stein CM, Chung CP, Daugherty JR, Smalley WE, Ray WA. An automated database case definition for serious bleeding related to oral anticoagulant use. Pharmacoepidemiol Drug Saf 2011 Jun;20(6):560-566. [doi: 10.1002/pds.2109] [Medline: 21387461]
10. Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. Thromb Res 2006;118(2):253-262. [doi: 10.1016/j.thromres.2005.06.015] [Medline: 16081144]
11. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. Stroke 2002 Oct;33(10):2465-2470. [doi: 10.1161/01.str.0000032240.28636.bd] [Medline: 12364739]
12. Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 2005 Apr;3(4):692-694. [doi: 10.1111/j.1538-7836.2005.01204.x] [Medline: 15842354]
13. Hauk L. Bleeding with anticoagulants: ACC releases expert consensus decision pathway to guide management. Am Fam Physician 2018 Jul 1;98(1):57-58. [Medline: 30215949]
14. Tomaselli GF, Mahaffey KW, Cuker A, et al. 2020 ACC expert consensus decision pathway on management of bleeding in patients on oral anticoagulants: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 2020 Aug 4;76(5):594-622. [doi: 10.1016/j.jacc.2020.04.053] [Medline: 32680646]
15. Hartenstein A, Abdelgawwad K, Kleinjung F, Privitera S, Viethen T, Vaitsiakhovich T. Identification of International Society on Thrombosis and Haemostasis major and clinically relevant non-major bleed events from electronic health records: a novel algorithm to enhance data utilisation from real-world sources. Int J Popul Data Sci 2023;8(1):2144. [doi: 10.23889/ijpds.v8i1.2144] [Medline: 38414540]
16. Liu H, Bielinski SJ, Sohn S, et al. An information extraction framework for cohort identification using electronic health records. AMIA Jt Summits Transl Sci Proc 2013;2013:149-153. [Medline: 24303255]
17. Settles B. Active learning literature survey. : University of Wisconsin–Madison; 2009 URL: https://minds.wisc.edu/server/api/core/bitstreams/8a78cf83-0702-4157-aeff-3cacb62f9ad5/content [accessed 2025-03-10]
18. proportion_ci. SciPy. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats._result_classes.BinomTestResult.proportion_ci.html#r67c9937f472a-3 [accessed 2026-04-01]
19. Lam KFY, Gopal V, Qian J. Confidence intervals for the F1 score: a comparison of four methods. arXiv. Preprint posted online on Sep 26, 2023. [doi: 10.48550/ARXIV.2309.14621]
20. An J, Hartenstein A, Choi SK, et al. Validation of electronic health record-based International Society on Thrombosis and Haemostasis bleeding algorithms in a US health system. Res Pract Thromb Haemost 2026 Apr;10(4):106616. [doi: 10.1016/j.rpth.2026.106616] [Medline: 42232248]
21. Gaspar F, Zayene M, Coumau C, et al. Natural language processing and ICD-10 coding for detecting bleeding events in discharge summaries: comparative cross-sectional study. JMIR Med Inform 2025 Aug 29;13:e67837. [doi: 10.2196/67837] [Medline: 40882207]
22. Zheng NS, Keloth VK, You K, et al. Detection of gastrointestinal bleeding with large language models to aid quality improvement and appropriate reimbursement. Gastroenterology 2025 Jan;168(1):111-120. [doi: 10.1053/j.gastro.2024.09.014] [Medline: 39304088]
23. Coumau C, Gaspar F, Zayene M, et al. Detection of antithrombotic-related bleeding in older inpatients: multicenter retrospective study using structured and unstructured electronic health record data. J Med Internet Res 2026 Jan 29;28:e77809. [doi: 10.2196/77809] [Medline: 41610431]
24. Lee HJ, Jiang M, Wu Y, et al. A comparative study of different methods for automatic identification of clopidogrel-induced bleedings in electronic health records. AMIA Jt Summits Transl Sci Proc 2017;2017(185–192):185-192. [Medline: 28815128]
25. Taggart M, Chapman WW, Steinberg BA, et al. Comparison of 2 natural language processing methods for identification of bleeding among critically ill patients. JAMA Netw Open 2018 Oct 5;1(6):e183451. [doi: 10.1001/jamanetworkopen.2018.3451] [Medline: 30646240]
26. Shung D, Tsay C, Laine L, et al. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. J Gastroenterol Hepatol 2021 Jun;36(6):1590-1597. [doi: 10.1111/jgh.15313] [Medline: 33105045]
27. Pedersen JS, Laursen MS, Rajeeth Savarimuthu T, et al. Deep learning detects and visualizes bleeding events in electronic health records. Res Pract Thromb Haemost 2021 May;5(4):e12505. [doi: 10.1002/rth2.12505] [Medline: 34013150]
28. Li R, Hu B, Liu F, et al. Detection of bleeding events in electronic health record notes using convolutional neural network models enhanced with recurrent neural network autoencoders: deep learning approach. JMIR Med Inform 2019 Feb 8;7(1):e10788. [doi: 10.2196/10788] [Medline: 30735140]
29. Wang X, Huang L, Xu S, Lu K. How does a generative large language model perform on domain-specific information extraction?—A comparison between GPT-4 and a rule-based method on band gap extraction. J Chem Inf Model 2024 Oct 28;64(20):7895-7904. [doi: 10.1021/acs.jcim.4c00882] [Medline: 39375999]
30. Patel PV, Davis C, Ralbovsky A, et al. Large language models outperform traditional natural language processing methods in extracting patient-reported outcomes in inflammatory bowel disease. Gastro Hep Adv 2024;4(2):100563. [doi: 10.1016/j.gastha.2024.10.003] [Medline: 39877865]

XSL•FO
RenderX

31. Lybarger K, Yetisgen M, Uzuner Ö. The 2022 n2c2/UW shared task on extracting social determinants of health. J Am Med Inform Assoc 2023 Jul 19;30(8):1367-1378. [doi: 10.1093/jamia/ocad012] [Medline: 36795066]
32. Patra BG, Lepow LA, Kasi Reddy Jagadeesh Kumar P, et al. Extracting social support and social isolation information from clinical psychiatry notes: comparing a rule-based natural language processing system and a large language model. J Am Med Inform Assoc 2025 Jan 1;32(1):218-226. [doi: 10.1093/jamia/ocae260] [Medline: 39423850]

Abbreviations

CNS: central nervous system
Dx Code: diagnosis code
ED: emergency department
EHR: electronic health record
FDA: Food and Drug Administration
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
INR: International Normalized Ratio
ISTH: International Society on Thrombosis and Haemostasis
LLM: large language model
MIMIC-III: Medical Information Mart for Intensive Care
NLP: natural language processing
OHNLP: Open Health Natural Language Processing
PRBC: packed red blood cells

Edited by A Coristine; submitted 09.Nov.2025; peer-reviewed by MA Le-Pogam, S Ananiadou; revised version received 23.Sep.2026; accepted 24.Sep.2026; published 20.Nov.2026.

Please cite as:
Draper EW, Anderson JA, Shah S, Brown L, Hanson MR, Negley KD, Lutz JM, Anderson AE, O'Meara JG, Jakowenko ND, McBane RD
Comparison of Rule-Based Natural Language Processing and Diagnosis Code Approaches to Detecting Bleeding in the Electronic Health Record: Algorithm Validation Within a Retrospective Cohort Study
JMIR Cardio 2026;10:e87424
URL: https://cardio.jmir.org/2026/1/e87424
doi:10.2196/87424

© Evan W Draper, Jan A Anderson, Surbhi Shah, Leilia Brown, Margaret R Hanson, Kristin D Negley, Jennifer M Lutz, Andrea E Anderson, John G O'Meara, Nicholas D Jakowenko, Robert D McBane. Originally published in JMIR Cardio (https://cardio.jmir.org), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cardio, is properly cited. The complete bibliographic information, a link to the original publication on https://cardio.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX