# Contents

## Original Papers

**Implantation of an Innovative Intracardiac Microcomputer System for Web-Based Real-Time Monitoring of Heart Failure: Usability and Patients’ Attitudes (e21055)**  
Giuseppe D’Ancona, Monica Murero, Sebastian Feickert, Hilmi Kaplan, Alper Öner, Jasmin Ortak, Hueseyin Ince. .......................................................... 2

**Patient Perspectives With Telehealth Visits in Cardiology During COVID-19: Online Patient Survey Study (e25074)**  
Aniruddha Singh, Natalie Mountjoy, Doug McElroy, Shilpi Mittal, Bashar Al Hemyari, Nicholas Coffey, Kristen Miller, Kenneth Gaines. ........................... 28

**Predicting Cardiovascular Risk Using Social Media Data: Performance Evaluation of Machine-Learning Models (e24473)**  
Anietie Andy, Sharath Guntuku, Srinath Adusumalli, David Asch, Peter Groeneveld, Lyle Ungar, Raina Merchant. ................................................................. 41

**Heart Rate Measurements in Patients with Obstructive Sleep Apnea and Atrial Fibrillation: Prospective Pilot Study Assessing Apple Watch’s Agreement With Telemetry Data (e18050)**  
Pauline Huynh, Rongzi Shan, Ngozi Osuji, Jie Ding, Nino Isakadze, Francoise Marvel, Garima Sharma, Seth Martin. ................................................................. 47

**Development and Validation of an Automated Algorithm to Detect Atrial Fibrillation Within Stored Intensive Care Unit Continuous Electrocardiographic Data: Observational Study (e18840)**  
Allan Walkey, Syed Bashar, Md Hossain, Eric Ding, Daniella Albuquerque, Michael Winter, Ki Chon, David McManus. ................................................................. 56

**Internet-Delivered Exposure-Based Therapy for Symptom Preoccupation in Atrial Fibrillation: Uncontrolled Pilot Trial (e24524)**  
Josefin Särnholm, Helga Skuladóttir, Christian Rück, Sofia Klavebäck, Eva Ólafsdóttir, Susanne Pedersen, Frieder Braunschweig, Brjann Ljótsson. ........................................ 66

**Digital Footprint of Academic Vascular Surgeons in the Southern United States on Physician Rating Websites: Cross-sectional Evaluation Study (e22975)**  
Qi Yan, Katherine Jensen, Rose Thomas, Alyssa Field, Zheng Jiang, Christian Goei, Mark Davies. ................................................................. 79

## Review

**User Engagement With Smartphone Apps and Cardiovascular Disease Risk Factor Outcomes: Systematic Review (e18834)**  
Erin Spaulding, Francoise Marvel, Rebecca Piasecki, Seth Martin, Jerilyn Allen. ................................................................. 13
Implantation of an Innovative Intracardiac Microcomputer System for Web-Based Real-Time Monitoring of Heart Failure: Usability and Patients’ Attitudes

Giuseppe D’Ancona¹, MD, PhD; Monica Murero², PhD; Sebastian Feickert¹, MD; Hilmi Kaplan¹, MD; Alper Öner³, MD, PhD; Jasmin Ortak³, MD, PhD; Hueseyin Ince¹,³, MD, PhD

¹Department of Cardiology, Vivantes Hospital Am Urban, Berlin, Germany
²Federico II University, Naples, Italy
³Department of Cardiology, Rostock University, Rostock, Germany

Corresponding Author:
Giuseppe D’Ancona, MD, PhD
Department of Cardiology
Vivantes Hospital Am Urban
Dieffenbachstraße 1
Berlin, 10967
Germany
Phone: 49 (030) 130 22 5
Email: rgea@hotmail.com

Abstract

Background: Heart failure (HF) management guided by the measurement of intracardiac and pulmonary pressure values obtained through innovative permanent intracardiac microsensors has been recently proposed as a valid strategy to individualize treatment and anticipate hemodynamic destabilization. These sensors have potential to reduce patient hospitalization rates and optimize quality of life.

Objective: The aim of this study was to evaluate the usability and patients’ attitudes toward a new permanent intracardiac device implanted to remotely monitor left intra-atrial pressures (V-LAP, Vectorious Medical Technologies, Tel Aviv, Israel) in patients with chronic HF.

Methods: The V-LAP system is a miniaturized sensor implanted percutaneously across the interatrial septum. The system communicates wirelessly with a “companion device” (a wearable belt) that is placed on the patient’s chest at the time of acquisition/transmission of left heart pressure measurements. At first follow-up after implantation, the patients and health care providers were asked to fill out a questionnaire on the usability of the system, ease in performing the various required tasks (data acquisition and transmission), and overall satisfaction. Replies to the questions were mainly given using a 5-point Likert scale (1: very poor, 2: poor, 3: average, 4: good, 5: excellent). Further patient follow-ups were performed at 3, 6, and 12 months.

Results: Use and acceptance of the first 14 patients receiving the V-LAP technology worldwide and related health care providers have been analyzed to date. No periprocedural morbidity/mortality was observed. Before discharge, a tailored educational session was performed after device implantation with the patients and their health care providers. At the first follow-up, the mean score for overall comfort in technology use was 3.7 (SD 1.2) with 93% (13/14) of patients succeeding in applying and operating the system independently. For health care providers, the mean score for overall ease and comfort in use of the technology was 4.2 (SD 0.8). No significant differences were found between the patients’ and health care providers’ replies to the questionnaires. There was a general trend for higher scores in patients’ usability reports at later follow-ups, in which the score related to overall comfort with using the technology increased from 3.0 (SD 1.4) to 4.0 (SD 0.7) (P=.40) and comfort with wearing and adjusting the measuring thoracic belt increased from 2.8 (SD 1.0) to 4.2 (SD 0.4) (P=.02).

Conclusions: Despite the gravity of their HF pathology and the complexity of their comorbid profile, patients are comfortable in using the V-LAP technology and, in the majority of cases, they can correctly and consistently acquire and transmit hemodynamic data. Although the overall patient/care provider satisfaction with the V-LAP system seems to be acceptable, improvements can be achieved after ameliorating the design of the measuring tools.

Trial Registration: ClinicalTrials.gov NCT03775161; https://clinicaltrials.gov/ct2/show/NCT03775161
Introduction
Heart failure (HF) is a pandemic with important public health implications [1,2]. Patient management guided by the measurement of intracardiac and pulmonary pressure values, obtained through innovative permanent intracardiac microsensors, has been recently proposed as a valid strategy to individualize and anticipate the management of patients with chronic HF, with the goal of reducing their hospitalization rate and optimizing their quality of life [3-6]. In this context, the patients’ perspective on the use and acceptance of these innovative implantable technologies has been poorly studied.

We here report our experience with implantation of a new intracardiac device designed to monitor the left intra-atrial pressure (LAP) of patients with chronic HF through an internet-based information system. The applicability and effectiveness of this technology are currently under evaluation in a multicenter prospective trial (V-LAP study). We here focus on evaluation of device usability and satisfaction as perceived by both patients and health care providers.

Methods
Study Design
This study was developed as part of a multicenter prospective study (ClinicalTrials.gov NCT03775161) aimed at assessing the safety, usability, and performance of an intracardiac microsensor (V-LAP) implanted in patients with chronic HF that are subject to multiple rehospitalizations for acute decompensation. The trial was reviewed and approved by the ethical and scientific committees of the participating centers. All patients recruited signed an informed consent form to the processing and use of data for research purposes.

System
The V-LAP system (Vectorious Medical Technologies, Tel Aviv, Israel) is the latest-generation system that enables monitoring of the patient LAP. The left atrium is the left heart chamber located directly above the left ventricle, which is the portion of the heart mainly involved in HF. The pressure inside the left atrium accurately reflects the changes in pressure within the left ventricle and can therefore be used to monitor cardiac function changes during the different phases of HF. The V-LAP system is a miniaturized sensor that is implanted completely percutaneously (ie, without incision) from the femoral vein (groin) and is anchored across the interatrial septum with the sensor portion protruding into the left atrium (Figure 1). The implant is a pressure microsensor with a low-profile design (<18 mm long and 3.9 mm in diameter) that allows for taking pressure measurements (Figure 2). The V-LAP sensory implant is fixed within the interatrial septum, usually on its thinnest area, the fossa ovalis. The implant is comprised of a hermetically sealed body that encases the sensing elements and electronics, and a nitinol braided anchor (Figure 1). The anchor has two discs, and when the implant is fully deployed, the distal and proximal discs are positioned on the left and right sides of the interatrial septum, respectively, whereas the implant body traverses the septum. The microsensor implanted inside the heart communicates wirelessly with an external system. The external system includes a lightweight, wearable, flexible sash-like loop (wearable belt companion device) that the patient can easily wear over clothing around the chest for 1-3 minutes daily (Figure 1 and Figure 3). This unit remotely powers the implant, interrogates it, and communicates LAP information to health care professionals at the HF clinic via a cellular gateway (Figure 1 and Figure 4). The external system can be used in the clinic or at any location.

After implantation, measurements are performed once or twice a day to precisely monitor the hemodynamic status of the patient.
Figure 1. Intracardiac V-LAP Vectorious device implanted on the left side of the interatrial septum and the cycle of use.

Figure 2. Modified low-profile thoracic belt.

Figure 3. External measuring device (companion device, thoracic belt) and measurement performed after “belt” wearing.

Figure 4. Gateway companion device.
Predischarge Education and Usability/Satisfaction Evaluation

Before implantation of the device, the patients were informed about the technology and the implanting procedure. After implantation of the intracardiac microsensor, patients were instructed on how to independently perform daily measurements of intracardiac pressure using the external system, chest belt, and the associated gateway. Detailed information and use training were carried out on the first day after implantation. The information/educational session lasted 60 minutes and involved the patient, their health care providers (nurses and doctors specialized in the diagnosis and treatment of HF), a home care provider (whenever available), and two product technicians from the sponsor.

The session was divided into three modules: (1) familiarization with the gateway (ie, the device for power supply and data transmission to the cloud), (2) familiarization with the measuring wireless belt, and (3) appropriateness of measurement position to guarantee good communication between the internal cardiac sensor and the external measuring unit for acquisition of measurements.

An official user manual, approved by the ethics committee, was made available to both the patients and health care providers. During the first follow-up visit, 1 month after implantation, the patients and health care providers were asked to fill out a questionnaire.

The questionnaire consisted of several structured questions focusing on the usability of the system, ease in performing the various required tasks (data acquisition and transmission), and overall satisfaction. Replies to the questions were mainly given using a 5-point Likert scale (1: very poor, 2: poor, 3: average, 4: good, 5: excellent). Patient questionnaires were performed at every follow-up visit (1, 3, 6, and 12 months).

Statistical Analysis

Data are presented as frequencies for categorical variables and as means (SD) for normally distributed continuous variables. Scores achieved in the questionnaire responses of patients and health care providers were compared using the $\chi^2$ test and Student $t$ test as appropriate (Likert scale results were compared considering the values as numerical) to test if the health condition and the patients’ comorbid profile impacted self-reported usability and satisfaction.

Patients’ responses at first and last follow-ups were also compared to document if usability and satisfaction improved over time. The SPSS Statistics 25 program was used for all analyses.

Results

Demographics and User Responses

This study focuses on the data obtained from the first 21 patients receiving the V-LAP implant worldwide (at the time of writing of this manuscript, a total of 22 patients have been treated with the implant worldwide). Table 1 shows the patient demographics and clinical data. No periprocedural complications or in-hospital mortality were observed. Usability questionnaires were completed at the time of the first follow-up visit (approximately 1 month) after discharge by the patients and their health care providers. Table 2 shows the specific questions included in the questionnaires and the scores for each question for the 14 patients and 15 health care providers. As the study is ongoing, follow-up data of the remaining patients are still being collected.
Table 1. Demographic and clinical profile of the first 21 patients.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD), range</td>
<td>67.0 (10.32), 49-86</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Body mass index (kg/m²), mean (SD), range</td>
<td>29.44 (3.41), 24.16-36.7</td>
</tr>
<tr>
<td>CRT/ICD, n (%)</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Creatinine (mg/dL), mean (SD), range (n=20)</td>
<td>1.55 (0.52), 0.88-2.6</td>
</tr>
<tr>
<td>eGFR (mL/min/1.73m²), mean (SD), range (n=19)</td>
<td>54.35 (20.15), 24.0-90.7</td>
</tr>
<tr>
<td>Hemoglobin (g/dL), mean (SD), range (n=20)</td>
<td>13.53 (1.73), 10.6-16.7</td>
</tr>
<tr>
<td>6-Minute walk (meters), mean (SD), range (n=17)</td>
<td>221.38 (139.15), 27.5-450.0</td>
</tr>
<tr>
<td>Saturation O₂ (%), mean (SD), range (n=17)</td>
<td>96.41 (2.37), 92.0-100.0</td>
</tr>
<tr>
<td>LVEF (%) mean (SD), range (n=18)</td>
<td>30.78 (11.3), 15.0-55.0</td>
</tr>
<tr>
<td>Heart rate (beats/minute), mean (SD), range</td>
<td>72.81 (10.02), 55.0-97.0</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg), mean (SD), range</td>
<td>72.1 (9.63), 55.0-92.0</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg), mean (SD), range</td>
<td>115.52 (14.52), 90.0-147.0</td>
</tr>
<tr>
<td>Mean RAP (mmHg), mean (SD), range (n=14)</td>
<td>9.29 (7.12), 1.0-22.0</td>
</tr>
<tr>
<td>PASP (mmHg), mean (SD), range (n=17)</td>
<td>45.0 (15.53), 6.0-68.0</td>
</tr>
<tr>
<td>Mean PCWP (mmHg), mean (SD), range (n=16)</td>
<td>19.38 (7.32), 8.0-37.0</td>
</tr>
<tr>
<td>LAP invasive, mean (SD), range (n=14)</td>
<td>18.57 (7.8), 10.0-37.0</td>
</tr>
</tbody>
</table>

aCRT: cardiac resynchronization therapy.
bICD: intracardiac defibrillator.
cGFR: glomerular filtration rate.
dLVEF: left ventricular ejection fraction.
eRAP: right atrial pressure.
fPASP: pulmonary artery systolic pressure.
gPCWP: pulmonary capillary wedge pressure.
hLAP: left atrial pressure.
Table 2. Usability follow-up questionnaires completed by 14 patients and 15 health care providers.

<table>
<thead>
<tr>
<th>Question</th>
<th>Patients score(^a)</th>
<th>Health care providers score</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success in applying and operating the system, n (%)</td>
<td>13 (92.9)</td>
<td>15 (100)</td>
<td>.40</td>
</tr>
<tr>
<td>Ease of wearing and fastening the belt, mean (SD)</td>
<td>3.7 (1.2)</td>
<td>4.2 (0.8)</td>
<td>.20</td>
</tr>
<tr>
<td>Ease of holding the belt at the appropriate measurement position, mean (SD)</td>
<td>3.7 (1.2)</td>
<td>4.2 (0.8)</td>
<td>.20</td>
</tr>
<tr>
<td>Ease of measurement initiation(^b), mean (SD)</td>
<td>4.2 (1.3)</td>
<td>N/A(^c)</td>
<td>N/A</td>
</tr>
<tr>
<td>Level of comfort during measurement(^b), mean (SD)</td>
<td>4.0 (1.1)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Level of clarity of when the measurement is finished, mean (SD)</td>
<td>4.3 (1.1)</td>
<td>4.8 (0.4)</td>
<td>.10</td>
</tr>
<tr>
<td>Ease of unlocking the belt at the end of measurement(^b), mean (SD)</td>
<td>4.3 (1.1)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Level of clarity of when the belt needs to be charged(^b), mean (SD)</td>
<td>4.3 (1.2)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Ease of connecting the belt to the charger(^b), mean (SD)</td>
<td>4.2 (1.4)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Overall comfort and ease of use with the system, mean (SD)</td>
<td>3.9 (1.1)</td>
<td>4.2 (0.8)</td>
<td>.30</td>
</tr>
</tbody>
</table>

\(^a\)Scores for all questions except for success in applying and operating the system were measured on a 5-point Likert scale (1: very poor, 2: poor, 3: average, 4: good, 5: excellent) 1 month after implantation.

\(^b\)Only included in the patient questionnaire.

\(^c\)N/A: not applicable.

Patient Questionnaire

Figure 5 summarizes the questionnaire scores at the first follow-up visit for the first 14 patients. The overall comfort in use of the technology achieved a mean score of 3.9 at first follow-up (1 month), with 93% of the patients succeeding in applying and operating the system consistently and independently (Table 2). The lowest score was found for the ease in wearing, locking, and holding the measuring unit (belt) at the predetermined appropriate measurement position to guarantee good communication between the internal cardiac sensor and the external measuring unit (Table 2). Patients seemed to be comfortable in starting the measurements, during the measurements, and understanding when the measurement was completed and that the belt had to be unlocked before reconnecting to the charger (Table 2 and Figure 5).
Health Care Providers Questionnaire

All 15 health care providers included in this analysis were able to apply and operate the technology (Table 2). The mean score for overall ease and comfort in use of the technology, and for the ease in making the patient wear the thoracic belt and placing it in the appropriate measuring position reached 4.2 (Table 2). No significant differences were found between the patients’ and health care providers’ replies to the questionnaires (Table 2).

Table 3. Patient questionnaire responses at first and last follow-up.

<table>
<thead>
<tr>
<th>Question</th>
<th>First follow-up, mean (SD)</th>
<th>Last follow-up, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wearing and locking the belt</td>
<td>2.8 (1.0)</td>
<td>4.2 (0.4)</td>
<td>.02</td>
</tr>
<tr>
<td>Holding the belt for measurement</td>
<td>3.2 (1.6)</td>
<td>3.8 (0.8)</td>
<td>.50</td>
</tr>
<tr>
<td>Starting the measurement</td>
<td>3.0 (1.5)</td>
<td>4.4 (1.3)</td>
<td>.50</td>
</tr>
<tr>
<td>Comfort during measurements</td>
<td>3.2 (1.3)</td>
<td>3.8 (1.0)</td>
<td>.20</td>
</tr>
<tr>
<td>Ending measurement</td>
<td>3.6 (1.5)</td>
<td>4.2 (0.8)</td>
<td>.60</td>
</tr>
<tr>
<td>Unlocking the belt</td>
<td>3.4 (1.5)</td>
<td>4.2 (0.8)</td>
<td>.60</td>
</tr>
<tr>
<td>Charging and signal interpretation</td>
<td>3.2 (1.7)</td>
<td>3.7 (1.2)</td>
<td>.70</td>
</tr>
<tr>
<td>Connection to the charger</td>
<td>3.2 (2.0)</td>
<td>4.2 (1.3)</td>
<td>.40</td>
</tr>
<tr>
<td>Overall comfort</td>
<td>3.0 (1.4)</td>
<td>4.0 (0.8)</td>
<td>.40</td>
</tr>
</tbody>
</table>

Patient Questionnaire at First and Last Follow-Up

Table 3 summarizes the patients’ responses for each of the nine key questions at first and last follow-up. There was a general trend for higher scores of usability during follow-up, with an increase in the score for overall comfort with using the technology and specifically with wearing and adjusting the measuring thoracic belt.

Discussion

Principal Results and Comparison with Prior Work

The annual incidence of HF is increasing rapidly with an estimated worldwide prevalence of over 37.7 million people [1,2]. In its chronic phase, HF is the result of a functional cardiac insufficiency that can have multiple etiologies and manifests with numerous symptoms that compromise the patient’s quality of life. The most common symptoms of HF are shortness of breath (dyspnea), poor exercise tolerance (asthenia), and fluid
retention (edema). HF is associated with significant morbidity and mortality, and confers a substantial burden on health systems in the industrialized world. Indeed, HF is the leading cause of hospitalization among adults and the elderly.

To date, the treatment of chronic HF remains predominantly reactive, focusing on drug adaptation once signs and symptoms of HF exacerbation occur. Despite continuous improvements in the long-term management of HF, patients still experience acute exacerbation of this chronic disease, resulting in recurrent hospitalization. Therefore, current HF management strategies remain inefficient in tackling hospital readmissions, and in containing morbidity and mortality. Randomized controlled trials investigating the use of external wearable technologies designed to remotely monitor patients with HF have failed to demonstrate a clear reduction in hospitalization rates [3]. The failure of these technologies may be due to the limits of the biometric parameters measured and transmitted for patient management. In fact, the telemedicine systems most widely adopted in HF management make use of sensor-based wearable devices for measuring body parameters that normally change only in a later phase of HF exacerbation. For example, changes in body temperature (as result of altered peripheral perfusion), tissue impedance (resulting from subcutaneous tissue water content), body weight (as a consequence of water retention), and urinary production (as a consequence of reduced renal perfusion) are all delayed markers of HF exacerbation.

Given the inability of noninvasively accumulated data to help in preventing hospitalizations, it has become necessary to make a paradigm shift in the use of chronic HF management strategies. As part of this shift, it is essential to integrate innovative information and communication technologies that can identify early precursors of the forthcoming exacerbation of stable HF, even if an invasive microcomputer implantation procedure may be required. Clinical evidence shows that pulmonary and intracardiac pressure increases for up to several weeks before the onset of decompensated HF symptoms. Although the CHAMPION trial showed efficacy of an implantable pulmonary artery pressure sensor to manage HF patients at risk for rehospitalizations [4], having direct measurements of left heart pressure adds sensitivity for patients affected by HF and with additional cardiac conditions [3,5,6]. In particular, the use of permanent intracardiac microsensors can detect changes in cardiac function accurately and in advance of exacerbation requiring rehospitalization. In this way, appropriate treatment can be optimized and carried out quickly, anticipating the onset of symptoms.

Correct and early acquisition of intracardiac pressure can justify prompt interventions to be undertaken at preventing hospitalizations for exacerbation of HF. Implantable intracardiac sensors allow for the remote acquisition, measurement, and analysis of patients’ meaningful data in real time. Although these technologies, as well as data derived from their use obtained to date, have generated and will continue to generate broad attention, their effectiveness and application in everyday life will depend on adequate acceptance and adoption by the treated patients. Despite substantial effort in developing and optimizing these devices, understanding the treated patients’ perspective and perception is crucial to guarantee the smooth and constant application of these costly technologies, as well as their further improvement. To the best of our knowledge, this study is the first to dedicate specific attention to patient satisfaction and ease of use after implantation of an intracardiac device for HF monitoring. For this reason, specific comparison with previous literature cannot be performed.

The main finding emerging from this study is that, despite the gravity of their HF pathology and the complexity of their comorbid profile, patients are comfortable in using the V-LAP technology and, in the majority of cases, they can correctly and consistently acquire and transmit hemodynamic data. It must be noted that before inclusion in the trial and implantation of the V-LAP technology, patients had been adequately selected, evaluating not only their clinical profile but also their psychological status, along with their attitudes toward the disease and the possible medical and behavioral measures to be undertaken to reduce hospitalization, morbidity, and mortality. In this context, our findings cannot be generalized to the plethora of patients affected by chronic HF, many of whom have difficulties in using mobile devices or performing even the simplest of daily activities. Moreover, a patient-tailored educational session was provided after device implantation with the participation of all present and future actors involved in patient management. The educational session was structured to train patients and their respective health care providers on the use of the technology and to test the correct application of the taught modules during the days of hospitalization after device implantation. Although future technological improvements will possibly lead to simplification of the patient/health care provider tasks, the continuous and direct involvement and support of the devices’ manufacturing companies should be envisaged. This can further guarantee adequate education and training of an increasing number of treated patients and of the health care providers involved in their management.

Because perceived ease of use is one of the most important factors that can increase the adoption of mobile health systems [7], a critical appraisal should be given to our findings. Despite the overall comfort in adopting the V-LAP technology, interviewed patients and respective health care providers reported the lowest scores when assessing the ease in wearing and fastening the thoracic belt and in consistently finding its appropriate position for ideal measurements. Multiple iterations are performed during hospitalization and before discharge to determine the thoracic belt’s most appropriate position allowing for optimal wireless/radiofrequency communication with the intracardiac sensor. Once the best position is identified, a picture is taken that is given to both the patient and health care provider as reference for future measurements. Frequent changes of the heart position within the chest cavity may be necessary due to physiologic and pathologic variations in cardiac hemodynamics and geometries, particularly in patients affected by HF. These variations will reflect upon the position of the intracardiac sensor, and consequently upon the wireless interaction between the external belt and intracardiac sensor, ultimately influencing the eventual sequence of signal transmission/detection.

Based on our findings, firmware version improvements have been developed and implemented by the sponsor, and a new mechanical design of the thoracic belt will be available in the
very near future. The new design of the external system takes into consideration the patients’ challenges in securely fastening the belt connector to allow for continuous and uninterrupted communication with the intracardiac implant.

Moreover, and most importantly, a major improvement that is currently in development will resolve the challenges encountered in reproducing the exact positioning of the belt around the patient’s chest to guarantee adequate communication with the intracardiac implant. The new design of the companion belt has a smaller profile that will enhance placement around the patient’s chest, and allow for intuitive and precise placement in the predetermined position (Figure 3).

Adequate involvement of health care providers is crucial to guarantee the success of newly introduced and innovative technologies for monitoring HF patients. In this context, health care–related wearable and implantable technologies alone will not have the desired effect on patients’ health status improvement. In fact, data collected from these devices need to be interpreted and used within previously structured frameworks, allowing for solid and continuous interactions among patients and health care providers [3]. Interestingly, in spite of possible differences in age, health status, and digital/technology literacy between patients and health care providers, we did not find any significant difference in ease and comfort of use of the V-LAP technology. Although we are aware of the tremendous improvements already achieved with the V-LAP technology to treat HF, we do believe that a few challenges need to be overcome with the aim of further minimizing the path for data collection, transmission, analysis, and therapy adjustments. In particular, we foresee that the next generation of intracardiac monitoring systems should allow for hemodynamic and clinical data to be collected automatically without the need for actual measurements to be taken by the patient. As clinicians, we envision the possibility of an autopowered intracardiac sensor that will independently detect and transmit information about the patient’s hemodynamic status while performing their daily activities. In this light, emerging automatized systems of mobile health management based on the upcoming Internet of Things capabilities should be envisaged to increase the number of potential users accessing state-of-the-art technology. Although the average age of the currently most affected patients concerns a generation that is often technologically illiterate, it seems realistic that in only a few years, the coming generations of HF patients will find the handling of digital medical devices a matter of course and an uncomplicated task.

Finally, although beyond the scope of this study, we can confirm that V-LAP technology has immediate clinical applicability by supporting actual HF therapy changes. After observing variations in the patients’ intracardiac hemodynamic measurements taken from home, therapy adjustments were promptly updated by the remote health care providers to reflect the real-time condition, thereby avoiding hospitalization. This has been particularly useful during the ongoing COVID-19 pandemic, reducing the risk of contagion of these very fragile patients during their travel to the hospital or within the premises of outpatient clinics [8,9].

At present, patients do not receive direct feedback about the value of their own collected intracardiac data. In the near future, a companion app elaborating body-sensing data through artificial intelligence and machine learning systems may inform not only health care providers but also patients and their caregivers by offering information about the patient’s health status, proposing customized psychological comfort, and sending notifications and reminders aimed at the optimization of HF therapy. In fact, automated protocols based on artificial intelligence and machine learning are already transforming the management of other chronic diseases such as diabetes mellitus [10] and could eventually be used to maximize the potential of the V-LAP technology in the automated treatment of HF patients. In this context, it should also be emphasized that although the V-LAP system is currently mainly used to detect LAPs, correct and automated interpretation of the recorded LAP curves will facilitate the real-time monitoring of additional cardiac parameters such as the heart rhythm and mitral valve function.

Limitations

The main limitation of this study is the small number of patients involved at present. It should be kept into consideration that the discussed technology has been only very recently introduced and is still under evaluation in a clinical trial. In fact, the patients analyzed in this study represent the majority of all patients receiving the V-LAP device worldwide.

Moreover, as emphasized above, results in terms of usability and adoption of the technology may be biased by the adequate selection of patients as part of the trial’s inclusion and exclusion protocol, which involved evaluating their psychological status, attitude toward the disease and its management, and their desire for being involved with this innovative technology. Finally, because the primary and secondary objectives of the trial were not usability and satisfaction, the sample could not be adequately sized to draw definitive conclusions on these two matters.

Conclusions

Despite the gravity of their HF pathology and the complexity of their comorbid profile, patients are comfortable in using the V-LAP technology and, in the majority of cases, they can correctly and consistently acquire and transmit hemodynamic data. The overall patient/care provider satisfaction with the V-LAP system seems to be high. The scores of patients and respective health care providers were in the range of average to good with respect to assessing the ease in performing simple but crucial tasks such as wearing and fastening the thoracic belt, and more specifically in consistently finding its appropriate position for ideal measurements. Improvements in the external thoracic belt design have been very recently introduced and will hopefully further optimize patients’ and health care providers’ acceptance and adoption of this technology.
Acknowledgments
The trial is partially supported by a European Community funding program (Horizon 2020-European Commission).

Conflicts of Interest
Vectorious Medical Technologies has supported the V-LAP trial and provided a grant to cover expenses related to the publication of this manuscript. GD has received consultancy fees from Vectorious Medical Technologies. The other authors have no conflicts of interest to declare.

References

Abbreviations

HF: heart failure
LAP: left intra-atrial pressure
User Engagement With Smartphone Apps and Cardiovascular Disease Risk Factor Outcomes: Systematic Review

Erin M Spaulding¹,², PhD, BSN; Francoise A Marvel³,⁴, MD; Rebecca J Piasecki¹, BSN; Seth S Martin²,³,⁴,⁵,⁶, MD, MHS; Jerilyn K Allen¹,⁴,⁶, SCD, MS, FAAN

¹Johns Hopkins University School of Nursing, Baltimore, MD, United States
²The Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States
³Digital 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
⁴Johns Hopkins University School of Medicine, Baltimore, MD, United States
⁵Johns Hopkins University Whiting School of Engineering, Baltimore, MD, United States
⁶Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Erin M Spaulding, PhD, BSN
Johns Hopkins University School of Nursing
525 North Wolfe Street
Baltimore, MD, 21205
United States
Phone: 1 603 724 0604
Email: espauld2@jhu.edu

Abstract

Background: The use of mobile health (mHealth) interventions, including smartphone apps, for the prevention of cardiovascular disease (CVD) has demonstrated mixed results for obesity, hypercholesterolemia, diabetes, and hypertension management. A major factor attributing to the variation in mHealth study results may be mHealth user engagement.

Objective: This systematic review aims to determine if user engagement with smartphone apps for the prevention and management of CVD is associated with improved CVD health behavior change and risk factor outcomes.

Methods: We conducted a comprehensive search of PubMed, CINAHL, and Embase databases from 2007 to 2020. Studies were eligible if they assessed whether user engagement with a smartphone app used by an individual to manage his or her CVD risk factors was associated with the CVD health behavior change or risk factor outcomes. For eligible studies, data were extracted on study and sample characteristics, intervention description, app user engagement measures, and the relationship between app user engagement and the CVD risk factor outcomes. App user engagement was operationalized as general usage (e.g., number of log-ins or usage days per week) or self-monitoring within the app (e.g., total number of entries made in the app). The quality of the studies was assessed.

Results: Of the 24 included studies, 17 used a randomized controlled trial design, 4 used a retrospective analysis, and 3 used a single-arm pre- and posttest design. Sample sizes ranged from 55 to 324,649 adults, with 19 studies recruiting participants from a community setting. Most of the studies assessed weight loss interventions, with 6 addressing additional CVD risk factors, including diabetes, sleep, stress, and alcohol consumption. Most of the studies that assessed the relationship between user engagement and reduction in weight (9/13, 69%), BMI (3/4, 75%), body fat percentage (1/2, 50%), waist circumference (2/3, 67%), and hemoglobin A₁c (3/5, 60%) found statistically significant results, indicating that greater app user engagement was associated with better outcomes. Of 5 studies, 3 (60%) found a statistically significant relationship between higher user engagement and an increase in objectively measured physical activity. The studies assessing the relationship between user engagement and dietary and diabetes self-care behaviors, blood pressure, and lipid panel components did not find statistically significant results.

Conclusions: Increased app user engagement for prevention and management of CVD may be associated with improved weight and BMI; however, only a few studies assessed other outcomes, limiting the evidence beyond this. Additional studies are needed to assess user engagement with smartphone apps targeting other important CVD risk factors, including dietary behaviors, hypercholesterolemia, diabetes, and hypertension. Further research is needed to assess mHealth user engagement in both inpatient...
and outpatient settings to determine the effect of integrating mHealth interventions into the existing clinical workflow and on CVD outcomes.

*(JMIR Cardio 2021;5(1):e18834) doi: 10.2196/18834*

**KEYWORDS**

mHealth; smartphone; mobile phone; engagement; cardiovascular disease; health behaviors; risk factors

**Introduction**

**Background**

Heart disease remains the leading cause of death in the United States [1]. In 2011, the American Heart Association (AHA) set a strategic impact goal of decreasing deaths from cardiovascular diseases (CVDs) and stroke by 20% by 2020; thus, efforts have been made to improve health behavior and reduce the prevalence of risk factors for heart disease, including smoking, overweight and obesity, physical inactivity, poor nutrition, diabetes mellitus, hypercholesterolemia, and hypertension [1]. According to the 2015 National Health Interview Survey, 79% of US adults reported not achieving adequate physical activity (PA) [1]. According to the 2014 National Health and Nutrition Examination Survey, 46% of US adults have hypertension (based on the 2017 American College of Cardiology and AHA guidelines), 40% of US adults have moderately elevated or high total cholesterol, and 38% of US adults are obese [1]. Although still important risk factors for CVD, the prevalence of smoking and diagnosed diabetes is 15% and 9%, respectively, among US adults, which are lower than the other risk factors [1]. However, the prevalence of diabetes is increasing, whereas the prevalence of smoking among adults is decreasing [1].

Mobile health (mHealth), “the use of mobile computing and communication technologies, [such as smartphone applications]...for health services and information,” is an innovative approach that could be a potentially effective means of involving individuals in health promotion and CVD management [2]. Although the prevalence of smartphone ownership among adults in the United States is high [3-6], including 73% smartphone ownership among individuals with CVD risk [7], a state-of-the-science article demonstrated that there are conflicting findings about the effectiveness of mHealth interventions for CVD prevention in improving CVD health behavior change and risk factor outcomes, such as weight management, PA, hypertension management, diabetes management, and lipid control [2]. One potential but important cause of conflicting results may be differential user engagement with the interventions. However, directly assessing the relationship between user engagement with smartphone apps and CVD health behavior change and risk factor outcomes was not a goal of that review.

Engagement with (smartphone app) interventions is considered a precondition for effectiveness and is of particular concern for behavior change interventions [8]. Although the field of user engagement with health interventions is in the early stages, there is work that has been conducted to reach a consensus on how best to conceptualize and operationalize user engagement with these interventions [8-10]. For this systematic review, user engagement with smartphone apps is conceptually defined as the “emotional, cognitive, and behavioral experience of a user with a [smartphone application] that exists, at any point in time and over time, [to varying degrees]” [11]. User engagement is a dynamic process that likely coincides with behavior change to ultimately improve health outcomes [8]. Yardley et al [8] proposed one potential model, including 4 phases, of this process. In phase 1, individuals would engage with the smartphone app and prepare for behavior change [8,12]. In phase 2, individuals would partake in behavior change, mediated by sustained user engagement [8,12]. In phase 3, individuals would continue to partake in behavior change but may disengage from the smartphone app if no longer needed to sustain behavior change [8,12]. Finally, in phase 4, individuals may re-engage with the smartphone app if there is a lapse in behavior change [8,12]. Unlike other behavioral change interventions, mHealth interventions allow for the assessment of user engagement or intervention fidelity. If user engagement is determined to be associated with CVD health behavior change and risk factor outcomes, clinicians and providers could use summary reports of user engagement with mHealth interventions as a proxy for determining how individuals adhere to health behavior recommendations.

The process of user engagement with mHealth interventions can be measured either subjectively via self-report (eg, focus groups, observation, think-aloud activities, ecological monetary assessments, interviews, and questionnaires) to capture the emotional and cognitive experiences or objectively via physiological measurements or app analytics (eg, ecological monetary assessments, eye tracking, time spent on a page, revisits to app) to capture the behavioral experiences [9-11]. When measuring user engagement with smartphone apps, the more relevant measures include focus groups, interviews, questionnaires, ecological monetary assessments, and app analytics. App analytics measure the behavioral manifestations of user engagement with smartphone apps and can be divided into intersession and intrasession measures [11]. Intersession measures assess long-term user engagement with smartphone apps across multiple sessions [11]. Intrasession measures assess user engagement with smartphone apps within a single session [11]. For this systematic review, the emotional and cognitive aspects of user engagement are operationalized through questionnaires and the behavioral aspects of user engagement through app analytics.

There has been considerable research across multiple disciplines, examining what factors are associated with higher user engagement, including intervention content (feedback, goal setting, reminders, self-monitoring, and social support features), modes of content delivery (control, credibility, novelty, personalization, and professional support features), demographic characteristics (age, computer literacy, education, ethnicity,
employment, and gender), and psychological characteristics (experience of well-being, mental health, motivation, and self-efficacy) [10]. However, a key question to address is whether the degree of user engagement with an mHealth intervention correlates with achieving the targeted outcomes, in this case, CVD risk factor modification and outcomes. Determining whether user engagement with smartphone apps is associated with improved CVD health behavior change and risk factor modification will be critical for determining their clinical utility in the future.

**Objectives**

As smartphone ownership becomes more prevalent and individuals increasingly use their devices for health information and management, with 62% of smartphone owners found in a prior study to have used their smartphone in the past year to look up health information [6], we require a better understanding of user engagement with smartphone apps. To our knowledge, no systematic reviews have been previously conducted with the primary aim of assessing the relationship between user engagement with smartphone apps and CVD health behavior change and risk factor outcomes. Schoeppe et al [13] conducted a systematic review to evaluate the efficacy of interventions that used smartphone apps to improve diet, PA, and sedentary behavior. However, they only examined the relationship between user engagement and improvements in PA and healthy eating and cited the need for additional research to examine the relationship between user engagement and the outcomes of interest [13]. Therefore, the purpose of this systematic review, conducted 4 years following the work of Schoeppe et al [13], is to determine if user engagement with smartphone apps for the prevention and management of CVD is associated with improved CVD health behavior change and risk factor outcomes.

**Methods**

**Search Strategy and Eligibility Criteria**

ES searched PubMed, EBSCOhost, and CINAHL for articles published in English between 2007 and 2020. The review was limited to this period, as smartphones were not available until 2007. No search limitations were placed on participant age, setting, or population. Specific limitations were not placed on the population to identify individuals enrolled along the spectrum of CVD prevention (primordial, primary, and secondary). Although no search restrictions were placed on study duration, there were restrictions placed on study design that aimed to return studies conducted using a correlational design or nested within a randomized controlled trial (RCT), quasi-experimental design, or mixed methods design.

The following keywords were used to identify candidate studies: (Disease Management OR Disease Prevention OR Obesity OR Overweight OR Weight Loss OR Heart Diseases OR Vascular Diseases OR Cardiovascular Diseases OR Coronary Artery Diseases OR Heart Failure OR Hypertension OR Diabetes OR Exercise OR Physical Activity) AND (Mobile Applications OR mHealth OR Mobile Health OR iPhone OR Android OR Smartphone) AND (Engage* OR Experience* OR Usage OR Usability OR Involv*) AND (Randomized Controlled Trial OR Non-Randomized Controlled Trial OR Evaluation Studies OR Quasi-Experimental OR Mixed-methods OR Correlation Studies). The final searches were conducted on January 28, 2020. The complete search strategy can be found in Multimedia Appendix 1. The search terms around user engagement are meant to encompass both objective and subjective experiences. Unlike the other CVD risk factors discussed previously, we decided not to include smoking in the search strategy, as its prevalence is low and decreasing among US adults. Covidence, a software for managing and streamlining the systematic review process, was used to screen the returned studies and remove duplicates.

Studies were eligible for inclusion if they (1) evaluated user engagement with a smartphone app; (2) included a smartphone app that was used by an individual to manage his or her cardiovascular health; (3) assessed CVD health behavior change or risk factor outcomes (ie, medication adherence; symptom management; and changes in diet, PA, weight, and biomarkers) for primordial, primary, or secondary prevention of CVD (hypertension, coronary artery disease, obesity, diabetes, myocardial infarction, or heart failure), not including stroke; and (4) assessed whether user engagement with a smartphone app was associated with the CVD health behavior change or risk factor outcome.

Studies were excluded if the sample size for the mHealth intervention group was less than 50 participants to reduce the likelihood of drawing conclusions from insufficiently powered studies. Although there were no specific search limitations placed on the study population, with the intention of identifying individuals enrolled along the spectrum of CVD prevention (primordial, primary, and secondary), if the study was not focused on CVD management or prevention, they were excluded. Examples of study populations excluded for this reason included those where there may have been elements of the intervention that aimed to improve cardiovascular health, but overall the focus was on improving psychological distress, chronic kidney disease management, type 1 diabetes management, stroke recovery, or fertility among couples attempting conception. The reasons for exclusion were assigned based on a specific hierarchical structure, moving from broader to more specific exclusion criteria. Coauthors progressed sequentially through the following reasons for exclusion: (1) intervention or population not related to CVD behavior change, (2) smartphone app was not used by the patient, (3) less than 50 participants in the mHealth intervention groups or with engagement data, (4) no measure of user engagement, (5) outcome not related to CVD prevention or management, and (6) relationship between user engagement and CVD outcome not assessed. Once an article met any of these exclusion criteria, the coauthors assigned that as the reason for exclusion and did not continue to assess the article for the subsequent exclusion criteria.

**Screening Process and Data Extraction**

Each retrieved title and abstract were screened by ES to determine eligibility to qualify for full-text review. Occasional full-text reviews were completed when the operationalization of user engagement with the smartphone app was unclear from
the abstract. The articles identified for full-text review were independently examined for inclusion by ES and JA. A consensus was reached on all the articles eligible for inclusion, and a third reviewer was not needed. ES extracted data on study and sample characteristics, intervention description, measures of user engagement with the smartphone app, and results regarding the relationship between user engagement with the smartphone app and the CVD health behavior change or risk factor outcome. In some instances, ES reviewed associated protocol papers to obtain the necessary data on study design, quality, and intervention description for the included articles. Data were extracted into a table to summarize the findings for the narrative results of this review.

In this review, if multiple treatment arms in a study used a smartphone app, the results with regard to the relationship between user engagement and the outcome of interest are presented as they were in the original study (ie, either combined or separated user engagement metrics across groups). If the relationship between user engagement and the health outcome was reported at multiple time points, the end of the treatment time point was used by default when determining if app user engagement was or was not significantly associated with the outcome of interest. In the Results section, we characterize a statistically significant association found in the desired direction between user engagement and the outcome of interest (ie, a positive association between user engagement and PA or a negative association between user engagement and weight) as a positive finding. We also label a nonstatistically significant association in the expected direction, no association, or an association in the opposite direction than expected as a negative finding. Findings that ended up not being reported were also labeled as a negative finding. ES and RP independently assessed the methodological rigor of the included studies via the Joanna Briggs Checklist for analytical cross-sectional studies [14]. A consensus was reached between these 2 authors on the methodological rigor for each study, and a third reviewer was not required. The checklist for analytical cross-sectional studies was chosen, as the focus for this review was on the relationship between user engagement and the CVD health behavior change or risk factor outcome of interest, not the difference in the outcome between intervention groups or the change in the outcome over time (ie, the correlational designs nested within the RCT or quasi-experimental designs).

Results

Results of the Search

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines provided the structure for the flow of articles throughout the critical review process, which is shown in Figure 1 [15]. A total of 1964 records were identified from the 3 electronic databases. Then, 546 duplicate records were removed, and the titles and abstracts of the remaining 1418 records were reviewed. Of the 1418 records, 155 were identified for full-text review. Of 155 articles, 131 were excluded, and 24 articles (ie, studies), assessing 22 individual interventions, were deemed eligible for inclusion in this systematic review. Of the 24 studies included in this review, 16 (67%) were published in mHealth or technology journals, and 8 (33%) were published in medical or clinical journals. These findings likely reflect journal preferences, with medical and clinical journals, perhaps favoring studies that focus more on clinical outcomes and not necessarily user engagement with mHealth interventions.
Study Characteristics

The study characteristics are summarized in Multimedia Appendix 2 [16-39]. Of the 24 studies, 17 used an RCT design [16-32], 4 used a retrospective analysis [33-36], and 3 used a single-group pre-and-posttest quasi-experimental design [37-39]. Although 20 studies used an RCT or single-group pre-and-posttest quasi-experimental design, it was the correlational studies nested within these larger parent studies that were of interest for this review. For the 17 RCTs, intervention duration ranged from 6 weeks [22] to 24 months [17,24]. For the 3 single-group pre- and-posttest quasi-experimental studies, intervention duration ranged from 3 months [37] to 6 months [38,39]. For the studies using a prospective design, follow-up ranged from 1 month [25,26,28] to 24 months [17,24]. All 24 studies were conducted in developed countries, including the United States (n=11), Australia (n=3), Canada (n=2), Spain (n=2), the United Kingdom (n=1), Norway (n=1), Finland (n=1), Japan (n=1), Singapore (n=1), and Korea (n=1). Overall, 2 studies with the same first author assessed one intervention (MyFitnessPal) [25,26], and another 2 studies with the same first author also assessed one intervention (Noom) [33,34]. Therefore, although there were 24 studies, only 22 individual interventions were assessed.
Of the 17 RCT studies, 14 conducted an a priori power analysis. Of these 14 RCTs, 3 detected a significant difference in the primary outcome between groups [21,23,27], 9 did not detect a significant difference in the primary outcome between groups [16,17,22,24-26,29,31,32], and 2 did not report the results of the difference in the primary outcome between groups [18,30]. One of the studies that did not detect a significant difference between groups did not achieve the target sample size [32]. Three RCTs did not conduct an a priori power analysis [19,20,28], one of which was a pilot study [28]. However, one study still detected a significant difference in the primary outcome between groups [19]. Among the RCTs, only one study conducted a post hoc power analysis [37-39]. One study had a very large sample size and found a significant difference in the primary outcome over time [37]. The other 2 were pilot studies and did not conduct inferential statistics regarding the primary outcome [38,39]. None of the 4 retrospective studies conducted a power analysis [33-36]. One study detected a significant difference in the primary outcome over time [36]; however, the other 3 studies did not report differences in the primary outcome over time [33-35].

Overall sample sizes ranged from 55 [38] to 324,649 participants [35]. Although this review was not limited to a specific age group, all eligible studies comprised adults, but one study enrolled participants as young as 16 years [28]. All 24 studies reported on sex, with 16 of the samples consisting largely of women (range 51%-91%) [22,38], 7 of the samples consisting largely of men (range 51%-100%) [16,21], and one split evenly [39]. Thirteen studies reported sample races [17-19,21-28,32,39], with 7 consisting largely of White participants [18,19,22,24-26,28]. Seventeen studies reported on sample educational level, with the majority of the participants within each individual study having attended some college [16,26,28-32,38,39]. Ten studies limited their sample to participants who were overweight or obese [17,19,21-26,28,35], and another 11 studies reported enrolling participants with a baseline mean or majority percentage BMI indicating overweight or obesity [16,20,22,27,29-31,33,34,38,39]. The majority of the studies recruited participants from a community setting, except for 5 studies that recruited participants from hospitals [21], a hospital-based diabetes management education program [32], primary care [23,29], and a community health care facility [39].

All 22 interventions assessed included an app for addressing CVD risk factors. Of 22 interventions, 11 (50%) included commercial apps [19,20,22,25-27,32-37,39] and 11 (50%) included investigator-developed apps [16,18,21,23,24,28-31,38]. In addition, 73% (16/22) of the interventions consisted of not only an app but also other components, including in-person meetings, emails, text messaging, phone consults, websites, Facebook groups, blogs, and podcasts [16-26,29-31,36,38,39]. Of 22 interventions, 6 (27%) consisted solely of an app [27,28,32-35,37]. Twelve apps were also paired with tracking devices (ie, pedometers, weight scales, glucose meters) [18,20-24,29-31,35,38,39]. Most of the studies assessed weight loss interventions, with 27% (6/22) of the interventions addressing additional CVD risk factors, including type 2 diabetes, sleep, stress, excessive alcohol consumption, and smoking [18,20,31,32,36,39]. The interventions focused on weight loss used a variety of strategies, including nutrition and PA tracking-energy expenditure, diet and exercise education, podcasts, social support, recipes, and behavior change techniques.

### User Engagement With Smartphone App Measures

The behavioral manifestations of user engagement with smartphone apps were assessed using app analytics. Only intersession measures were used in the included studies. In 3 of the 17 studies that combined other intervention strategies with the app (ie, participants had the choice to use both the app and other intervention strategies such as a website), they did not differentiate app from website user engagement and provided a combined effect [16,17,20]. The intersession measures used were general usage and overall self-monitoring over a defined period. The general usage of the interventions was assessed in 11 studies through multiple means, including tracking the number of log-ins per week or month, number of usage days, total time spent using the app, and number of times the different app features were used [16,20,24,28-32,36-38]. In 2 studies, self-reported weekly app use was assessed as a proxy for intersession general usage measures [19,22]. In 14 of the studies, researchers assessed overall self-monitoring (eg, total number of days self-monitoring or entries made in the app) [16-18,21,23-27,32-35,39]. User engagement with the smartphone apps was also categorized by frequency and pattern of self-monitoring or general usage in 14 of the studies [17,19-21,23,27,29-31,35-39].

In 33% (8/24) of the studies, there was no indication of how frequently participants were able to or expected to engage with the intervention (ie, intended engagement) [18-20,28,31,32,35,36]. In 67% (16/24) of the studies, there was some indication of how frequently participants were able to or expected to engage with the intervention [16,17,21-27,29,30,33,34,37-39], whether it was broad instructions to self-monitor behaviors daily or almost daily [16,17,22,25,26,29,30,33,34,37,38], multiple times a day [21,27], or more specific recommendations for each feature in the intervention [23,24,39]. It was not always clear in the studies that indicated participants were able to self-monitor within the intervention daily if they were explicitly told or prompted to do so.

Overall, 6 studies included a smartphone app in more than one treatment arm [22,24-26,30,31]. In 2 of these studies, where there was no difference in the smartphone app between the 2 groups, user engagement data across both groups with the smartphone app were combined when assessing the relationship between user engagement and the outcome of interest [22,31]. In another study, where the versions of the smartphone app were different between the 2 arms (gamified app vs nongamified app), the app user engagement data were also combined across groups [29]. In studies assessing the MyFitnessPal app, user engagement was both combined and separated across groups when assessing the relationship between user engagement and the outcome of interest [25,26]. As the app did not differ across
groups, but rather instruction on how to engage with the app as well as supplementary material, in the narrative of this review, we present the combined results [25,26]. In another study, where both intervention arms had access to the same app, one without reminders and prompts, the results were presented separately for each arm [24]. Multimedia Appendix 2 provides a description of how user engagement with the smartphone apps was operationalized in each study. None of the studies reported on the subjective experience of user engagement [11] with a smartphone app; however, a few did report on satisfaction and usability of the app [9].

**Quality of Studies**

Of the 24 included studies, 21 used convenience or purposive sampling [16-28,30-32,35-39] and only 3 used random sampling [29,33,34], limiting the external validity of these results. Seventeen of the studies used an RCT design [16-32], one of which also used random sampling [29], strengthening the internal validity. In Multimedia Appendix 3, the methodological rigor for each study is presented using the Joanna Briggs Checklist for analytical cross-sectional studies [14]. The rules for how the studies were scored on methodological rigor for each question, directed by the Joanna Briggs Checklist for analytical cross-sectional studies guidelines, are also provided in Multimedia Appendix 3.

The questions that studies performed the worst on the Joanna Briggs Checklist for analytical cross-sectional studies were question 1 (“Were the criteria for inclusion in the sample clearly defined?”) in which 42% (10/24) of the studies received a score of yes, question 3 (“Was the exposure measured in a valid and reliable way?”) in which none of the studies received a score of yes, question 5 (“Were confounding factors identified?”) in which 42% (10/24) of the studies received a score of yes, and question 8 (“Was appropriate statistical analysis used?”) in which 38% (9/24) of the studies received a score of yes. The studies performed well on question 2 (“Were the study subjects and the setting described in detail?”) in which 71% (17/24) of the studies received a score of yes, question 4 (“Were objective, standard criteria used for measurement of the condition?”) in which 88% (21/24) of the studies received a score of yes, and question 7 (“Were the outcomes measured in a valid and reliable way?”) in which 88% (21/24) of the studies received a score of yes.

Of the studies that were not nested within an RCT, 5 found positive results [33-37], one demonstrated more mixed results [39], and one with a smaller sample size (n=55) found negative results [38]. Thus, it does not appear as though studies nested within a poorer quality design demonstrated more negative results. In general, the sample sizes were large (88% had a sample size ≥100 participants). Heterogeneous app analytics were used across studies to assess user engagement with smartphone apps, making it difficult to draw strong conclusions. In addition, in 3 studies, user engagement was operationalized across technology platforms, limiting the accurate assessment of the relationship between user engagement with the app and the outcome [16,17,20].

**Relationship Between User Engagement With Smartphone Apps and Health Outcome**

Multimedia Appendix 2 also provides a description of the relationship between user engagement with smartphone apps and the CVD health behavior change or risk factor outcome or both in each study. If the relationship between user engagement and the health outcome was reported at multiple time points, the end of the treatment time point was used by default when determining whether app user engagement was or was not significantly associated with the outcome of interest. There were, however, 2 studies that presented these findings at only a preliminary [32] or longer follow-up [29] time point; thus, the findings at these time points had to be used instead when making a determination.

**Changes inAnthropometrics**

Overall, 15 studies assessed the relationship between user engagement with a smartphone app and changes in anthropometrics, including weight, BMI, percent body fat, and waist circumference.

**User Engagement and Change in Weight**

Thirteen studies assessed the relationship between user engagement with a smartphone app and change in weight. Nine studies reported statistically significant greater weight loss with higher user engagement with a smartphone app, with follow-up ranging from 2 to 12 months [19-21,23,25-27,34,35]. Both general usage [19-21] and self-monitoring measures [23,25-27,34,35] of user engagement with a smartphone app were used. In particular, entering PA, dietary behaviors, and weight; more frequent upload of meal photographs; completing more educational articles; customizing more features; and posting on social platforms were significantly associated with greater weight loss [23,25-27,34,35]. However, simply commenting on other users’ posts was not [34]. Four studies that assessed the relationship between user engagement with a smartphone app and weight demonstrated negative results, with follow-up ranging from 3 to 24 months [17,24,28,39]. Self-monitoring measures [17,24,39] and general usage measures [24,28] of user engagement with smartphone apps were used. Although Lin et al [24] did not find statistically significant correlations between user engagement and weight change at 24 months (the length of intervention duration), at 6 months, both intervention arms with an app found that as app user engagement increased, weight decreased. In addition, at 12 months, this relationship remained statistically significant for the app intervention arm paired with group dietitian–led sessions and phone calls [24].

**User Engagement and Change in BMI**

Four studies assessed the relationship between user engagement with a smartphone app and BMI [18,20,21,33]. Both general usage [20,21] and self-monitoring measures [18,33] of user engagement with smartphone apps were used. Three of these studies reported a statistically significant greater reduction in BMI with higher user engagement with the smartphone app, with follow-up ranging from 6 to 12 months [20,21,33]. In particular, activities within the Noom weight loss app, including logging food and group participation (number of original posts
and comments and likes on others’ posts), were significantly associated with a reduction in BMI [33]. Among Gray Matters app users, there was no correlation between the average number of log-ins per day and change in BMI at 6 months [18].

**User Engagement and Change in Waist Circumference and Body Fat Percentage**

Three [20,21,27] and 2 [20,21] studies assessed the relationship between user engagement and waist circumference and body fat, respectively. In 2 of the studies, sustained users of the app and web technology interventions and more frequent meal photograph uploads were significantly associated with greater reduction in percent body fat [20] and waist circumference [20,27], with follow-up ranging from 2 to 12 months. One study that aimed to assess the relationship between user engagement and waist circumference and percent body fat at 6 months did not report the data [21].

**CVD Health Behavior Change**

Six studies assessed the relationship between user engagement with a smartphone app and change in PA or dietary behaviors or both [16,22,29,30,37,38], and one study assessed the relationship between app user engagement and diabetes self-care behaviors [32]. The changes in health behavior were collected via self-reported data entered into the app or through paired devices [16,22,29,30,37,38] or via surveys [16,29,30,32]. General usage [16,29,30,37,38] and self-monitoring measures [16,22] of user engagement with smartphone apps were used.

**User Engagement and Change in PA**

Among the 5 studies that assessed change in objectively measured PA (ie, step count and minutes of moderate to vigorous physical activity) [22,29,30,37,38], 3 studies found statistically significant associations between user engagement and increases in objectively measured PA with follow-up ranging from 3 months to 12 months [29,30,37]. Two of the 5 studies with smaller sample sizes (n=67 and n=55) and follow-up at 6 weeks and 6 months did not [22,38]. The relationship between user engagement with smartphone apps and self-reported PA was assessed across 3 studies with follow-up ranging from 3 to 12 months [16,29,30]. One study found that increased user engagement was associated with increased self-reported PA [30], and 2 studies did not [16,29]. In a study conducted by Edney et al [30], app engagement data were combined across multiple treatment arms using a gamified versus nongamified app. Although the results for the relationship between user engagement and step count and self-reported PA were not presented separately for each arm, they did report that gamified app users were more likely to be in the high user engagement group [30].

**User Engagement and Change in Dietary and Diabetes Self-Care Behaviors**

In 2 studies with follow-up ranging from 9 to 12 months, no general usage or self-monitoring measures of user engagement with the intervention were associated with a change in self-reported dietary behaviors [16,29]. One study that conducted an exploratory analysis assessing the relationship between overall app use and self-reported diabetes self-care behaviors at 3 months did not find a statistically significant association [32].

**Change in Risk Factors and Biomarkers**

Eight studies assessed the relationship between user engagement with a smartphone app and biomarkers, including hemoglobin A1c (HbA1c; n=5), heart rate (n=1), systolic blood pressure (SBP) and diastolic blood pressure (DBP; n=2), cholesterol (n=2), triglycerides (n=2), blood carotenoids (n=1), serum glucose (n=2), and insulin levels (n=1) [18,20,21,27,31,32,36,39]. General usage [18,20,31,32] and self-monitoring measures [18,21,27,32] of user engagement with smartphone apps were used.

**User Engagement and Change in HbA1c**

Five studies assessed the relationship between user engagement with a smartphone app and HbA1c [27,31,32,36,39]. Three of these studies found that general usage of the app overall was significantly associated with a decrease in HbA1c with follow-up ranging from 3 to 12 months [31,32,36]. However, when assessing the relationship between self-monitoring measures and decrease in HbA1c, the findings were slightly more mixed. One study found that more frequent meal photograph uploads, when comparing the highest tertile to the lowest tertile, were significantly associated with decreased HbA1c at 8 weeks [27], but 3 other studies found that meal or diet tracking was not associated with a decrease in HbA1c with follow-up ranging from 3 to 12 months [31,32,39]. In addition, greater use of the exercise features in one study was associated with a decrease in HbA1c at 3 months [32] but not in another study at 12 months [31]. Weight tracking [39], but not blood glucose tracking [31,32,39], was found to have a significant relationship with a decrease in HbA1c.

**User Engagement and Change in Blood Pressure, Lipid Panel, and Other Biomarkers**

Among the 3 other studies assessing the relationship between user engagement with a smartphone app and risk factors and biomarkers, one found mixed results [18] and the other 2 found negative results [20,21]. Among participants enrolled with the Gray Matters app (n=104), there was a statistically significant positive correlation between the average number of health behavior questions answered per day and improvement in total and high-density lipoprotein (HDL) cholesterol at 6 months [18]. However, the average number of logs completed per day was not significantly associated with resting heart rate, SBP, DBP, cholesterol, triglycerides, blood carotenoids, serum glucose, or insulin levels at 6 months [18]. Among users of an app and web-based technology intervention (n=118), there were no statistically significant relationships between sustained and nonsustained usage and a change in aerobic fitness (METmax), SBP, DBP, triglycerides, or total cholesterol at 12 months [20]. Among users of the SmartCare weight loss app, at 6 months, there were no statistically significant differences in lipid panel improvement (total cholesterol, HDL cholesterol, and triglycerides) between those who entered anthropometric data at least 3 times per week versus those who did so less than 3 times per week [21].
Table 1 provides the number of studies that looked at each outcome, the number of studies for each outcome that had a positive finding, and the number of studies for each outcome that had a negative finding. A positive finding refers to a statistically significant association found in the desired direction between user engagement and the outcome of interest (ie, a positive association between user engagement and step count or a negative association between user engagement and weight). A negative finding refers to a nonstatistically significant association in the expected direction, no association, or an association in the opposite direction than expected. None of the studies in this review found a significant association in the direction opposite to what was expected. For the study that ended up not reporting waist circumference and body fat percentage results [21], this was also categorized as a negative finding.

### Table 1. Findings for the relationship between user engagement and the cardiovascular disease health behavior change or health outcome.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies with positive finding, n (%)</th>
<th>Studies with negative finding, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (n=13)</td>
<td>9 (69)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>BMI (n=4)</td>
<td>3 (75)</td>
<td>1 (25)</td>
</tr>
<tr>
<td>Percent body fat (n=2)</td>
<td>1 (50)</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Waist circumference (n=3)</td>
<td>2 (67)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>Objectively measured physical activity (n=5)</td>
<td>3 (60)</td>
<td>2 (40)</td>
</tr>
<tr>
<td>Self-reported physical activity (n=3)</td>
<td>1 (33)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Self-reported diet (n=2)</td>
<td>0 (0)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Hemoglobin A1c(^a) (n=5)</td>
<td>3 (60)</td>
<td>2 (40)</td>
</tr>
<tr>
<td>Systolic blood pressure (n=2)</td>
<td>0 (0)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Diastolic blood pressure (n=2)</td>
<td>0 (0)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Total cholesterol (n=3)</td>
<td>0 (0)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>HDL(^b) cholesterol (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Triglycerides (n=3)</td>
<td>0 (0)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Resting heart rate (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Blood carotenoids (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Serum glucose (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Insulin levels (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>METmax (n=1)</td>
<td>0 (0)</td>
<td>1 (100)</td>
</tr>
</tbody>
</table>

\(^a\)HbA1c: hemoglobin A1c.  
\(^b\)HDL: high-density lipoprotein.

A meta-analysis was not conducted because the primary objective of this systematic review was not to assess the effectiveness of the interventions as a whole but rather, to determine whether increased app user engagement was associated with improvement in CVD health behavior change and risk factor outcomes. In addition, the heterogeneity in study designs and methods would have led to bias in a meta-analysis. However, as a sensitivity analysis, we assessed the relationship between user engagement with a stand-alone app versus an app plus other intervention components and weight, the most frequently assessed CVD health outcome. Of the 7 studies that assessed a stand-alone app [27,28,32-35,37], 4 assessed the relationship between user engagement and weight [27,28,34,35]. Of these 4 studies, 3 (75%) found positive results [27,34,35]. Of the 17 studies that assessed an app plus other intervention components, 9 assessed the relationship between user engagement and weight [17,19-21,23-26,39]. Of these 9 studies, 6 (67%) found positive results. Thus, in this sensitivity analysis, a higher percentage of studies assessing the relationship between user engagement with a stand-alone app and weight found positive results. However, the number of studies assessing a stand-alone app that also examined the relationship between user engagement and weight was small.

### Discussion

#### Principal Findings

We found 24 studies that met the eligibility criteria for this systematic review. This systematic review revealed that increased user engagement with a smartphone app, measured by general usage or self-monitoring entries or both, for the prevention and management of CVD may be associated with a reduction in weight (9/13 studies with positive findings) and BMI (3/4 studies with positive findings). Although only a few studies assessed the relationship between user engagement with a smartphone app and body fat percentage (1/2 studies with positive findings), waist circumference (2/3 studies with positive findings), and objectively measured PA (3/5 studies with positive findings), the findings were generally positive.
However, although only a few studies assessed the relationship between user engagement with a smartphone app and dietary (2 studies) and diabetes self-care behaviors (1 study), the results were all negative. Among the 5 studies that assessed the relationship between user engagement with a smartphone app and HbA1c, the results were promising (3/5 studies with positive findings). The 3 studies that assessed the relationship between user engagement with a smartphone app and the remaining biomarkers were largely statistically nonsignificant.

There are multiple explanations for why higher user engagement was associated with greater weight loss but not with reduction in biomarkers, such as blood pressure, total cholesterol, and triglycerides. First, the 3 studies that looked at these outcomes were likely underpowered to detect a significant relationship. In fact, Mattila et al [20] conducted post hoc analyses demonstrating that they were underpowered to detect significant associations between user engagement and change in SBP, DBP, total cholesterol, and triglycerides, although they were adequately powered to detect significant associations between user engagement and change in weight, BMI, waist circumference, and body fat percentage. The study by Hartin et al [18] was only powered to detect a medium effect size of 0.50 difference between the 2 treatment groups on SBP, DBP, total cholesterol, and triglycerides. The study by Oh et al [21] was powered to detect a 1.81 kg difference in weight between the 2 treatment groups, but they did not conduct power analyses for the other biomarker outcomes.

Second, only one of these 3 studies [21] incorporated a medication adherence component into the intervention. This intervention incorporated telephone consultations with nurses, exercise specialists, and clinical dietitians, which could include discussion regarding medications, as well as covering the costs of medications as an incentive to partake in the study [21]. However, it is unclear how frequently medication adherence was discussed as a part of these consultations. The other 2 studies did not include medication adherence components in their intervention [18,20]. It is possible that the lifestyle interventions in these 3 studies, largely focused on diet and PA, were not sufficient to significantly reduce these CVD biomarkers. In the future, researchers should consider adding education on CVD medications such as antihypertensives and statins as well as medication adherence self-monitoring capabilities into their mHealth interventions if they want to impact CVD biomarkers.

**Comparison With Prior Work**

Few systematic reviews have been conducted in this area, with most assessing the effectiveness of interventions on health outcomes or strategies to promote user engagement with mHealth interventions, as opposed to directly assessing the relationship between user engagement with mHealth and CVD health behavior change and risk factor outcomes. Semper et al [40] included 6 studies in their systematic review assessing the effectiveness of smartphone apps, which encourage dietary self-regulatory strategies, for weight loss among overweight and obese adults. This review demonstrated that participants using the smartphone apps in all studies lost at least some weight; however, when compared with other self-monitoring tools, there was no significant difference in the amount of weight loss [40]. Semper et al [40] did not report on user engagement with the apps. Although studies such as this have demonstrated that self-monitoring may be associated with greater weight loss, this does not provide a comprehensive picture of how user engagement with smartphone apps is associated with health behavior change and outcomes. Smartphone apps are capable of incorporating multiple behavior change strategies, such as goal setting, feedback, reminders, and social support features, and simply assessing self-monitoring of health behaviors may not be indicative of user engagement with the app as a whole. Thus, in this review, we build upon prior research by assessing user engagement with the app as a whole as well as individual features such as self-monitoring.

Schoeppe et al [13] conducted a systematic review to evaluate the efficacy of interventions that used smartphone apps to improve diet, PA, and sedentary behavior. Of the 23 included studies among adults, 17 demonstrated significant improvements in PA (n=13), diet (n=6), weight (n=4), blood pressure (n=2), sedentary behavior (n=1), fitness (n=1), and cholesterol (n=1) [13]. Eleven of the studies included in this review reported app usage to assess user engagement [13]. However, only 3 of these studies examined the association between app usage and changes in behavior and health outcomes, cautiously demonstrating that higher app usage was associated with improvements in PA and healthy eating [13]. The authors of this review recommended that further work be conducted to examine the relationship between user engagement and the outcomes of interest [13]. This systematic review fills this gap in the literature by building upon these findings and examining the relationship between user engagement and additional behavior change and health outcomes.

Another review assessed factors related to user engagement with internet behavioral interventions across many chronic health conditions, including type 2 diabetes, weight loss maintenance, and CVD [41]. They found that the interventions that adapted to individual needs, including record keeping, personalized feedback, and accountability, were more engaging [41]. In this systematic review, measures of user engagement with a smartphone app, such as self-monitoring (record keeping), were also frequently associated with improved risk factor outcomes, including a reduction in weight and BMI. Another systematic review of 14 RCTs looked at the effectiveness of technology-based strategies (eg, offering digital health intervention assistance) to promote engagement with digital health interventions (web-based platforms paired with emails, telephone calls, and texting) that targeted various health behaviors and conditions [42]. The studies often reported small-to-moderate effects of technology-based strategies on engagement compared with no strategy [42]. Previous reviews have focused on compiling strategies to promote user engagement with mHealth interventions. This systematic review builds upon previous work by assessing the relationship between user engagement with smartphone apps and the actual CVD health behavior change or risk factor outcome.
Limitations and Future Directions

The studies included in this review varied in how they operationalized and analyzed user engagement with smartphone apps, making it challenging to compare results across studies. None of the studies used intrasession measures of user engagement with a smartphone app, limiting this review to intersession measures. Intersession measures provide a better understanding of long-term user engagement with smartphone apps, but intrasession measures could provide valuable insight into how users engage within a single session. Researchers should consider using both types of measures. In addition, researchers should obtain both general usage and self-monitoring intersession measures. General usage measures will make it easier to compare results across studies, but self-monitoring activities within the app can provide more insight into specific intervention component use and whether engagement with them is associated with better outcomes.

In addition, 33% (8/24) of the studies in this review did not provide any indication of intended engagement with the intervention. Among the 67% (16/24) of the studies that did indicate how frequently participants were able to use certain features, it was not always clear whether participants were explicitly told or prompted on how frequently to engage. In the future, providing clear instructions for intended engagement and then determining whether not meeting, meeting, or exceeding intended engagement expectations are associated with achieving the intended outcomes will facilitate advancement in this field. Ultimately, seeking to determine what is considered sufficient user engagement with the intervention to achieve the intended outcomes (ie, effective engagement), as opposed to the current standard, that more user engagement is always better [8,10]. No studies have assessed the subjective experience of user engagement with the smartphone app, largely limiting this review to the behavioral manifestations of user engagement with the smartphone app via app analytics. This narrow focus on the behavioral aspects of user engagement is reflective of the state of the science but is not sufficient to assess the multidimensional concept of user engagement. Future studies could consider using the User Engagement Scale [43] or the eHealth Engagement Scale [44] to assess the subjective experience of user engagement.

The majority of the assessed apps focused on weight management. Additional studies are needed to assess user engagement with smartphone apps targeting hypercholesterolemia, diabetes, and hypertension, other important risk factors for CVD. In addition, most studies recruited participants from the community setting. In the future, more studies need to be conducted where participants are recruited in the inpatient or outpatient setting, and the apps are integrated with clinical care to determine whether this further affects the relationship between user engagement with smartphone apps and CVD health behavior change and risk factor outcomes.

There are also limitations specific to the conduction of this review, which should be taken into consideration. First, hand searching was not conducted as part of the search strategy. Second, only one author reviewed the titles and abstracts. However, if there was any question of whether a study should be included or excluded at this stage, it was progressed to full-text review, at which point 2 authors assessed the potentially eligible studies. Third, only one author extracted the data from the included studies to populate the table and results. Finally, it was outside the scope of the review to report on actual user engagement outcomes; however, this may be important for future reviews.

This systematic review also has many strengths. First, we searched multiple databases starting from when smartphones first became available. Second, 2 authors independently screened the full-text articles. Third, 2 authors independently assessed the quality of the included studies using a standardized assessment tool. Fourth, we reported the study findings on the relationship between both general usage and self-monitoring measures and the outcome of interest. Fifth, we provided clear delineation of the number of studies that had a positive or negative finding for each outcome via a table.

Conclusions

This systematic review found that user engagement with smartphone apps may be associated with risk factor outcomes, including reduction in weight and BMI, among adults using a smartphone app for CVD prevention and management. To draw stronger conclusions moving forward and to move toward the concept of effective engagement, the mHealth community needs to reach a consensus on how best to consistently operationalize user engagement with smartphone apps across multiple platforms and incorporate intended engagement with the intervention into measurement approaches.

Acknowledgments

ES received the following financial support for the research, authorship, and publication of this paper: NIH/NHLBI T32 HL007024 Postdoctoral Fellowship in Cardiovascular Epidemiology Institutional Training, NIH/NINR F31 NR017328 Ruth L. Kirschstein National Research Service Award, and NIH/NINR T32 NR012704 Pre-Doctoral Fellowship in Interdisciplinary Cardiovascular Health Research. RP has received the following financial support for the research, authorship, and publication of this paper: NIH/NINR F31 NR018362 Ruth L. Kirschstein National Research Service Award.

Conflicts of Interest

FM and SM are founders of and hold equity in the Corrie Digital Health platform. This arrangement has been reviewed and approved by Johns Hopkins University in accordance with its conflict of interest policies. The MiCORE study at Johns Hopkins (PI: SM) testing the Corrie Digital Health platform has received material support from Apple and iHealth and funding from the NIH/NINR F31 NR017328 Ruth L. Kirschstein National Research Service Award.
Maryland Innovation Initiative, Wallace H. Coulter Translational Research Partnership, Louis B. Thalheimer Fund, and Johns Hopkins Individualized Health Initiative. In addition to the above funding for the MiCORE study, SM has received research support from the American Heart Association (20SFRN35380046 and COVID19-811000), PCORI (ME-2019C1-15328), National Institutes of Health (P01 HL108800), Aetna Foundation, the David and June Trone Family Foundation, the Pollin Digital Innovation Fund, PJ Schafer Cardiovascular Research Fund, CASCADE FH, and Google. ES, RP, and JA have no conflicts of interest to declare.

Multimedia Appendix 1
Search strategy.
[DOCX File, 20 KB - cardio_v5ile1e18834_app1.docx ]

Multimedia Appendix 2
Study characteristics.
[DOCX File, 50 KB - cardio_v5ile1e18834_app2.docx ]

Multimedia Appendix 3
Quality assessment of studies included in this systematic review using the Joanna Briggs Checklist for analytical cross-sectional studies.
[DOCX File, 16 KB - cardio_v5ile1e18834_app3.docx ]

References


Abbreviations

AHA: American Heart Association

CVD: cardiovascular diseases

DBP: diastolic blood pressure

HbA1c: hemoglobin A1c

HDL: high-density lipoprotein

mHealth: mobile health

PA: physical activity

RCT: randomized controlled trial

SBP: systolic blood pressure

©Erin M Spaulding, Francoise A Marvel, Rebecca J Piasecki, Seth S Martin, Jerilyn K Allen. Originally published in JMIR Cardio (http://cardio.jmir.org), 03.02.2021. 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 http://cardio.jmir.org, as well as this copyright and license information must be included.
Patient Perspectives With Telehealth Visits in Cardiology During COVID-19: Online Patient Survey Study

Aniruddha Singh, MD; Natalie Mountjoy, PhD; Doug McElroy, PhD; Shilpi Mittal, MD; Bashar Al Hemyari, MD; Nicholas Coffey, MSc; Kristen Miller, BSc; Kenneth Gaines, MD

1Western Kentucky Heart and Lung Research Foundation, Bowling Green, KY, United States
2Western Kentucky University, Bowling Green, KY, United States
3Vanderbilt University Medical Center, Nashville, TN, United States
4University of Kentucky College of Medicine, Bowling Green, KY, United States

*these authors contributed equally

Abstract

Background: The rise of COVID-19 and the issue of a mandatory stay-at-home order in March 2020 led to the use of a direct-to-consumer model for cardiology telehealth in Kentucky. Kentucky has poor health outcomes and limited broadband connectivity. Given these and other practice-specific constraints, the region serves as a unique context to explore the efficacy of telehealth in cardiology.

Objective: This study aims to determine the limitations of telehealth accessibility, patient satisfaction with telehealth relative to in-person visits, and the perceived advantages and disadvantages to telehealth. Our intent was two-fold. First, we wanted to conduct a rapid postassessment of the mandated overhaul of the health care delivery system, focusing on a representative specialty field, and how it was affecting patients. Second, we intend to use our findings to make suggestions about the future application of a telehealth model in specialty fields such as cardiology.

Methods: We constructed an online survey in Qualtrics following the Patient Assessment of Communication During Telemedicine, a patient self-report questionnaire that has been previously developed and validated. We invited all patients who had a visit scheduled during the COVID-19 telehealth-only time frame to participate. Questions included factors for declining telehealth, patient satisfaction ratings of telehealth and in-person visits, and perceived advantages and disadvantages associated with telehealth. We also used electronic medical records to collect no-show data for in-person versus telehealth visits to check for nonresponse bias.

Results: A total of 224 respondents began our survey (11% of our sample of 2019 patients). Our recruitment rate was 86% (n=193) and our completion rate was 62% (n=120). The no-show rate for telehealth visits (345/2019, 17%) was nearly identical to the typical no-show rate for in-person appointments. Among the 32 respondents who declined a telehealth visit, 20 (63%) cited not being aware of their appointment as a primary factor, and 15 (47%) respondents cited their opinion that a telehealth appointment was not medically necessary as at least somewhat of a factor in their decision. Both in-person and telehealth were viewed favorably, but in-person was rated higher across all domains of patient satisfaction. The only significantly lower mean score for telehealth (3.7 vs 4.2, \( P=.007 \)) was in the clinical competence domain. Reduced travel time, lower visit wait time, and cost savings were seen as big advantages. Poor internet connectivity was rated as at least somewhat of a factor by 33.0% (35/106) of respondents.

Conclusions: This study takes advantage of the natural experiment provided by the COVID-19 pandemic to assess the efficacy of telehealth in cardiology. Patterns of satisfaction are consistent across modalities and show that telehealth appears to be a viable alternative to in-person appointments. However, we found evidence that scheduling of telehealth visits may be problematic and needs additional attention. Additionally, we include a note of caution that patient satisfaction with telehealth may be artificially inflated during COVID-19 due to external health concerns connected with in-person visits.
KEYWORDS
COVID-19; telehealth; cardiology; internet; broadband; patient satisfaction; restriction; survey

Introduction
In its most simplistic form, telehealth or telemedicine refers to the mixture of art and science to maintain health and prevent disease from a distance [1]. The definition of telehealth has evolved along with technological advances. Medicaid currently defines it as “two-way, real time interactive communication between the patient, and the physician or practitioner at the distant site. This electronic communication means the use of interactive telecommunications equipment that includes, at a minimum, audio and video equipment” [2,3]. The use of telehealth has been increasing, as demonstrated by the rise of telehealth visits among the commercially insured from 206 (0.02 per 1000) in 2005 to 202,374 (6.57 per 1000) in 2017. This annual growth rate of 52%, and the 261% increase between 2015-2017 alone, is likely associated with the rise of parity laws mandating coverage for such visits. The main contributors to this rise have been in primary care telehealth and tele–mental health visits [4]. The medium has been adopted by disciplines that require minimum physical exam findings, such as radiology and dermatology, while other, more heavily exam-dependent, specialties such as cardiology have been more resistant [5].

Perceived barriers from the physician-side include the lack of a comprehensive physical examination, technically challenged staff and patients, public resistance to telehealth, cost, reimbursement issues, and lower standards of care concerns [5,6]. Naser et al [7] conducted a literature review to present perspectives of telemedicine in cardiology in Bosnia and Herzegovina. This study provided an interesting take on how telemedicine is advancing in transitional countries and focused mainly on the different types of technology needed for a patient encounter. The key issue seemed to be development of software that would provide authentic data and be available for patients’ use. The authors suggested a primary limitation on the use of telemedicine, or information technology itself in medicine, was poor quality of software solutions and poor connectivity, with inadequate software maintenance. Although these and other technological factors can limit its use in rural areas, Naser et al [7] concluded that interactive video consultations provided better access to heart specialists and subspecialists than other means, accurate diagnosis, better treatment, reduction of mortality, and a significant reduction in costs.

Additionally, Di Lenarda et al [8] examined the strategic importance of innovative models of care for nonhospitalized patients with heart failure, along with the challenges and opportunities for its widespread clinical implementation. Their research revealed that technology development is mostly market driven, leading to an excess of data, unverifiable quality, and scarce utility. They recommended a multidisciplinary and multi-professional “Chronic Care Model” of integration between hospital and territory, and suggested that Italy’s active role in integrating telemedicine is helping to avoid heart failure hospitalizations.

Despite the continuing dialectic around the efficacy of telehealth in cardiology, the onset of the global COVID-19 pandemic necessitated a more or less immediate shift toward remote modalities to ensure continuation of care for cardiology patients, without increasing health risks. The transition has generated many important research questions about not only quality care but also patient use and perceptions of the novel modality. Will patients be able to access this care? Will they be satisfied with the experience? What are their perceived advantages and disadvantages to this new approach? Few studies have evaluated satisfaction with telemedicine in a broad range of cardiology patients, but what is available comes mostly from heart failure studies. Kraii et al [9] evaluated 14 publications from multiple databases. They found patients were satisfied with telemedicine but that the measurement of patient-reported outcomes, such as patient satisfaction with noninvasive telemedicine in patients with heart failure, is underexposed. None of the studies examined provided a clear definition or concept of patient satisfaction with telemedicine, and all studies evaluated patient satisfaction using different scales or questionnaires. The authors recommended that patient satisfaction become a more prominent theme in telemedicine research and that well-designed, validated, and standardized instruments with theoretic foundations were needed to measure patient satisfaction with telemedicine.

One such instrument, developed and validated by Agha et al [10], is a self-report questionnaire called the Patient Assessment of Communication During Telemedicine (PACT). The PACT is built on the four key domains of the physician-patient experience: patient-centered communication, clinical competence, interpersonal skills, and supportive environment [11-13]. The domain of “patient-centered communication” assesses the perception of the physicians’ active involvement with patients. Items regarding the “perceived clinical competence” of the physician focus on the patient’s experience with the clinical examination and their confidence in the physician’s clinical abilities. Patient perception of “interpersonal skills” includes patient’s emotional needs and comfort in discussing medical concerns with their providers. The “supportive environment” domain measures patients’ perception of professionalism with their cardiologist and other in-office personnel. The theoretical foundations of instruments like the PACT allow for a comparison between patients’ perceptions of telehealth visits and standard in-person visits; the four domains are transferable to both modalities.

Kentucky presently serves as an ideal study location in the United States for examining the efficacy of and patient satisfaction with telehealth in cardiology. In recent years, Kentucky has ranked in the top 10 states for prevalence of obesity (2018) and among the top five states for prevalence of diabetes (2016) [14,15]. These factors contributed to the state’s top 10 ranking in age-adjusted total cardiovascular deaths per 100,000 persons (from 2016 to 2018) [16]. This poor chronic health standing is compounded by the fact that Kentucky ranks
in the bottom 10 states for household income as of 2018 [17]. Economic constraints combined with rural geography contribute to a lack of internet availability; one of every four households in Kentucky lacks a broadband internet connection [18]. However, the rise of COVID-19 and issuance of a mandatory stay-at-home order for all nonessential employees by the Kentucky State government on March 16, 2020, necessitated use of a direct-to-consumer model for cardiology telehealth for adult patients.

The cardiovascular needs of Kentuckians, coupled with the limitations described, provides the context for a timely natural experiment. Here, we use a survey of cardiology patients to investigate the utility of telehealth from their perspective. Our primary objectives were to determine the existing limitations of telehealth accessibility, patient satisfaction with telehealth relative to traditional in-person visits in a situation where the mandatory shift to telehealth minimized self-selection bias, and the resulting perceived advantages and disadvantages to telehealth. Our intent was two-fold. First, we wanted to conduct a rapid postassessment of the mandated overhaul of the health care delivery system, focusing on a representative specialty field, and how it was affecting patients. We needed to know what was working and what was not so as to inform adaptive management in the near term. Second, we intended to use our findings to make suggestions about the future application of a telehealth model in specialty fields such as cardiology.

Methods

We employed a web-based survey and used existing electronic medical record (EMR) data to answer these research questions. Although an online survey may seem like an odd choice (the same barriers that may keep patients from using telehealth could also keep them from answering an online survey on a PC or other device, such as lack of broadband internet access or lack of computer skills), it afforded the rapid analyses required to answer these questions in real time.

Survey Sample

We intended to survey individuals who had appointments scheduled with their cardiologist at Western Kentucky Heart and Lung (WKHL) during the COVID-19 pandemic. WKHL is the primary cardiology and pulmonary and critical care training site for the University of Kentucky cardiovascular fellowship programs in Bowling Green, Kentucky and is associated with The Medical Center as its main hospital. WKHL office staff consolidated the contact information for all patients scheduled between March 15, 2020 (the start of telehealth-only appointments due to COVID-19), and the survey implementation date on June 7, 2020. The resulting pool consisted of 2019 patients across 7 cardiologists. Our research protocol and questionnaire were approved by the Institutional Review Board of the Medical Center (IRB #20-6-05-SinA-TeleCOVID). All respondents provided an informed consent and data were kept on a secure device.

We constructed the questionnaire using Qualtrics (Qualtrics International Inc) and sent a bulk invitation email with a direct link to the online questionnaire to all 2019 patients. We optimized the survey for mobile browsers and sent two reminders, both as text messages and emails, with a direct link to the questionnaire [19]. These reminders were sent after the first week and the day the survey closed. To increase participation, we informed invitees that we would donate to COVID-19 relief efforts at The Medical Center for each completed survey [20,21]. We also provided an assurance of confidentiality and included a statement of thanks to others that had responded in the reminder messages [22].

Survey Instrument

Data were collected via an anonymous online survey following Dillman et al’s [20] Tailored Design Method for internet surveys and included expert review by cardiologists using telehealth and was pilot-testing among 25 WKHL office staff and medical interns for validity. Our questionnaire closely followed the PACT, the patient self-report questionnaire developed and validated by Agha et al [10]. As with the PACT and other studies our questionnaire assessed perspectives across the four domains of patient satisfaction: patient-centered communication, perceived clinical competence, interpersonal skills, and a supportive environment [23,24]. Aside from a few additions to address the current context of COVID-19, the accessibility of specific telehealth modalities offered, and perceived advantages and disadvantages of telehealth, all questions and items were designed based on the PACT and other validated patient surveys regarding telehealth [25]. All questions, aside from the open response, race and ethnicity, and gender, required a response for the participant to continue. Respondents were not allowed to “go back” or review their answer choices at the end of the questionnaire. Excluding consent, the questionnaire was three pages long for patients who did not have a telehealth visit and four pages long for those who did. Each page had from 7 to 24 question items (in three blocks). Following consent, the survey began with demographic questions to ensure we could measure representation in our sample, especially because economic and health disparities may be related to demography as well as access to telehealth. Respondents were also asked if they had sought medical care during the pandemic, about their travel time to their cardiologist, and if they participated in telehealth through their cardiologist during the pandemic. The answer to this last question bifurcated respondents onto two different survey paths.

If a respondent answered “no” regarding their participation in telehealth, they were directed to a “No Tele” set of questions regarding potential barriers to their access of telehealth. They were asked what factors may have influenced their decision not to participate in a telehealth visit, which included not medically necessary, no access to a smartphone or other device, privacy concerns, preference for in-person visits, and an open response option to include other influential factors. Respondents were asked to rank each option on a 3-point Likert-type scale as not a factor, somewhat of a factor, or the primary factor.

If a respondent answered “yes” regarding their participation in telehealth, they were directed to a “Had Telehealth” set of questions. They were asked about the modality of their telehealth visit (eg, phone call or face-to-face with a smartphone, computer, or tablet) and which platform was used (eg, Zoom...
Correlations among survey items were computed using differences in ratings using chi-square tests of association. Differences among cardiologists versus in-person appointments was tested for significance using a Wilcoxon signed rank test. Differences among mean ratings for telehealth versus in-person appointments was tested for significance using a Kruskal-Wallis (KW) nonparametric test. Paired difference in the average ratings for telehealth visits (n=2019) and not suffering from nonresponse bias.

Following these two separate paths, all respondents concluded the survey with a section regarding perceptions of their standard in-person visits with their cardiologists. The first section asked respondents to rank their level of agreement, on a 5-point Likert type scale, with 11 positive statements regarding the four domains of patient satisfaction. Lastly, respondents were asked to rank their overall experience on a 5-point Likert-type smile scale [26].

Electronic Medical Record Data

Aside from data collection via the survey, we also used EMR data to determine the no-show rate for telehealth appointments during our research period as well as the standard no-show rate for in-person visits during the 10 weeks prior to the state stay-at-home order. These additional data were collected to help address our questions around access to care and to ensure our sample was representative (ie, that we received enough responses from those who declined or missed their telehealth visits) and not suffering from nonresponse bias.

Statistical Analysis

All statistical analyses were carried out using SYSTAT, version 13 (Systat Software Inc). Cronbach alpha was used to test for internal consistency and scale reliability among related questions. Paired difference in the average ratings for telehealth versus in-person appointments was tested for significance using a Wilcoxon signed rank test. Differences among cardiologists in mean ratings for telehealth versus in-person appointments were examined using Kruskal-Wallis (KW) nonparametric analysis of variance. Individual items were tested for significant differences in ratings using chi-square tests of association. Correlations among survey items were computed using Spearman rank correlations and interpreted for significance based on Bonferroni-adjusted criteria. Post hoc power analysis was used to determine the level of statistical power in our comparisons of satisfaction ratings between telehealth and in-person visits.

Results

Respondent Characteristics

A total of 224 unique individuals (based on Internet Protocol addresses) consented to take the survey (11% of our total sample of 2019). Of those, 86% (n=193) were recruited (ie, completed the first page and consented). Of those recruited, our completion rate was 62% (n=120), and early terminated surveys were analyzed by completed sections only. The vast majority of the 193 respondents identified as White, non-Hispanic (n=172, 89.1%); 10 (5.2%) respondents identified as African American, 2 (1.0%) as Hispanic/Latinx, 1 (0.5%) as Asian, and the remainder as unidentified; these percentages are consistent with the racial and ethnic diversity of the surrounding region [27]. The majority (n=190, 98.5%) described themselves as native English speakers. Respondents ranged from 18 to 100 years of age, with an average of 59.9 (SD 1.0) years. More than one-quarter (n=53, 27.5%) of individuals had sought medical care during the survey period; of these, slightly less than half (10.9%) did so for heart-related issues. Respondents reported a mean travel time to in-person appointments of nearly 40 minutes (mean 39.2, SE 2.5), with 9 (4.7%) indicating a 2- to 3-hour required commitment.

Access: Reasons for Declining Telehealth

Over the course of our study period, the no-show rate for scheduled telehealth appointments at WKHL was 17% (343/2019); the no-show rate of in-person visits in the 10 weeks prior to the switch to telehealth was also between 16% and 17% (526/3172). Among our 193 respondents, 28% (n=55) did not attend their scheduled telehealth visit. However, of the 32 respondents completing the section on barriers to telehealth, 20 (62.5%) indicated they did not realize they had been scheduled for a telehealth visit during the study time frame. There were 15 (47%) respondents that cited their opinion that a telehealth appointment was not medically necessary as at least somewhat of a factor in their decision; 20 (62.5%) cited a preference for in-person appointments as at least somewhat of a factor in their declining telehealth; 7 (21.9%) cited comfort with technology as playing a role in their decision, while a small percentage identified access to technology (n=2, 6.2%) or privacy concerns (n=2, 6.2%) as factors. These data are summarized in Table 1. Additional responses collected via open response included concerns about the validity of telehealth appointments to address cardiac conditions.
Table 1. Distribution of responses to survey items relating to respondents’ basis for opting out of telehealth and perceived advantages/disadvantages of telehealth by those who had a telehealth appointment.

<table>
<thead>
<tr>
<th>Survey items</th>
<th>No factor, n (%)</th>
<th>Somewhat, n (%)</th>
<th>Primary, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors in declining telehealth (n=32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not scheduled</td>
<td>5 (15.6)</td>
<td>7 (21.9)</td>
<td>20 (62.5)</td>
</tr>
<tr>
<td>Not medically necessary</td>
<td>17 (53.1)</td>
<td>5 (15.6)</td>
<td>10 (31.3)</td>
</tr>
<tr>
<td>Access to technology</td>
<td>30 (93.8)</td>
<td>0 (0.0)</td>
<td>2 (6.2)</td>
</tr>
<tr>
<td>Comfort with technology</td>
<td>25 (78.1)</td>
<td>5 (15.6)</td>
<td>2 (6.3)</td>
</tr>
<tr>
<td>Privacy concerns</td>
<td>30 (93.8)</td>
<td>1 (3.1)</td>
<td>1 (3.1)</td>
</tr>
<tr>
<td>Preference for in-person</td>
<td>12 (37.5)</td>
<td>13 (40.6)</td>
<td>7 (21.9)</td>
</tr>
<tr>
<td>Advantages to participating in telehealth (n=106)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced travel time</td>
<td>12 (11.3)</td>
<td>33 (31.1)</td>
<td>61 (57.5)</td>
</tr>
<tr>
<td>Reduced visit wait time</td>
<td>12 (11.3)</td>
<td>37 (34.9)</td>
<td>57 (53.8)</td>
</tr>
<tr>
<td>Travel or cost savings</td>
<td>19 (18.0)</td>
<td>44 (41.5)</td>
<td>43 (40.5)</td>
</tr>
<tr>
<td>Disadvantages to telehealth (n=106)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor internet connectivity</td>
<td>71 (67.0)</td>
<td>27 (25.5)</td>
<td>8 (7.5)</td>
</tr>
<tr>
<td>Device technology issues</td>
<td>82 (77.4)</td>
<td>19 (17.9)</td>
<td>5 (4.7)</td>
</tr>
<tr>
<td>Comfort with device/software</td>
<td>76 (71.7)</td>
<td>21 (19.8)</td>
<td>9 (8.5)</td>
</tr>
<tr>
<td>Communication issues</td>
<td>73 (68.9)</td>
<td>26 (24.5)</td>
<td>7 (6.6)</td>
</tr>
<tr>
<td>Privacy concerns</td>
<td>91 (85.8)</td>
<td>11 (10.4)</td>
<td>4 (3.8)</td>
</tr>
</tbody>
</table>

Patient Satisfaction: Telehealth Versus In-Person Visits

Both in-person and telehealth experiences were viewed favorably, but in-person more so. The highest ratings were seen on individual items relating to the cardiologist’s perceived competence, interpersonal skills, and interest in their patient’s medical concerns; this pattern was consistent across both telehealth and in-person formats. The lowest ratings were given on items relating to the cardiologist’s support for the patient’s emotions, perceived interest in establishing a medical partnership, and thoroughness of the clinical exam. Mean scores were nearly identical among three of the four survey domains, ranging between 4.32 and 4.33 out of 5. Only the clinical competence domain generated a lower mean score (4.23), and this was driven entirely by the low rating on the item related to the thoroughness of the clinical exam; when this item was excluded, the domain mean score improved to 4.33. There was also high reliability among items within each survey domain, as Cronbach alpha values ranged from .879 to .973. These data are summarized in Table 2.
Table 2. Summary of responses by those who participated in telehealth, characterizing their telehealth (n=106) and in-person (n=96) experiences.

<table>
<thead>
<tr>
<th>Survey domains, items, and mode</th>
<th>Strongly disagree, n (%)b</th>
<th>Disagree, n (%)b</th>
<th>Neither, n (%)b</th>
<th>Agree, n (%)b</th>
<th>Strongly agree, n (%)b</th>
<th>Mean (SE)b</th>
<th>r</th>
<th>P value b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient-centered communication</strong>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PCC-1. My cardiologist seemed interested in my medical concerns.</strong></td>
<td>5 (4.7)</td>
<td>1 (0.9)</td>
<td>8 (7.6)</td>
<td>30 (28.3)</td>
<td>62 (58.5)</td>
<td>4.35 (0.10)</td>
<td>0.46</td>
<td>.74</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>5 (5.2)</td>
<td>29 (30.2)</td>
<td>58 (60.4)</td>
<td>4.45 (0.09)</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>7 (6.6)</td>
<td>2 (1.9)</td>
<td>5 (4.7)</td>
<td>38 (35.8)</td>
<td>54 (51.0)</td>
<td>4.23 (0.11)</td>
<td>0.40</td>
<td>.22</td>
</tr>
<tr>
<td><strong>PCC-2. My cardiologist tried to find out everything that was concerning me.</strong></td>
<td>4 (3.8)</td>
<td>2 (1.9)</td>
<td>15 (14.1)</td>
<td>45 (42.5)</td>
<td>40 (37.7)</td>
<td>4.09 (0.09)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>1 (1.0)</td>
<td>7 (7.3)</td>
<td>34 (35.4)</td>
<td>52 (54.2)</td>
<td>4.39 (0.09)</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td><strong>PCC-3. My cardiologist was interested in establishing a medical partnership.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>8 (7.6)</td>
<td>44 (41.5)</td>
<td>47 (44.3)</td>
<td>4.20 (0.09)</td>
<td>0.40</td>
<td>.16</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>5 (5.2)</td>
<td>33 (34.3)</td>
<td>54 (56.3)</td>
<td>4.41 (0.09)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>7 (6.6)</td>
<td>2 (1.9)</td>
<td>5 (4.7)</td>
<td>38 (35.8)</td>
<td>54 (51.0)</td>
<td>4.23 (0.11)</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td><strong>PCC-4. Instructions and treatment plans were clear to me at the end of the visit.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>8 (7.6)</td>
<td>44 (41.5)</td>
<td>47 (44.3)</td>
<td>4.20 (0.09)</td>
<td>0.40</td>
<td>.54</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>5 (5.2)</td>
<td>33 (34.3)</td>
<td>54 (56.3)</td>
<td>4.41 (0.09)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>7 (6.6)</td>
<td>2 (1.9)</td>
<td>5 (4.7)</td>
<td>38 (35.8)</td>
<td>54 (51.0)</td>
<td>4.23 (0.11)</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical competence</strong>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CC-1. My cardiologist provided an appropriate level of medical care.</strong></td>
<td>5 (5.7)</td>
<td>2 (1.9)</td>
<td>8 (7.6)</td>
<td>39 (36.7)</td>
<td>51 (48.1)</td>
<td>4.20 (0.10)</td>
<td>0.43</td>
<td>.71</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>5 (5.2)</td>
<td>33 (34.4)</td>
<td>54 (56.2)</td>
<td>4.41 (0.09)</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>5 (4.7)</td>
<td>0 (0.0)</td>
<td>6 (5.7)</td>
<td>39 (36.8)</td>
<td>56 (52.8)</td>
<td>4.30 (0.09)</td>
<td>0.40</td>
<td>.007</td>
</tr>
<tr>
<td><strong>CC-2. My clinical exam was thorough.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>7 (6.6)</td>
<td>40 (37.7)</td>
<td>52 (49.1)</td>
<td>4.26 (0.10)</td>
<td>0.39</td>
<td>.27</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>3 (3.1)</td>
<td>6 (6.3)</td>
<td>27 (28.1)</td>
<td>58 (60.4)</td>
<td>4.42 (0.09)</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>7 (6.6)</td>
<td>0 (0.0)</td>
<td>6 (5.7)</td>
<td>37 (34.9)</td>
<td>59 (55.6)</td>
<td>4.38 (0.09)</td>
<td>0.42</td>
<td>.33</td>
</tr>
<tr>
<td><strong>CC-3. I had confidence in my cardiologist’s clinical competence.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>7 (6.6)</td>
<td>40 (37.7)</td>
<td>52 (49.1)</td>
<td>4.26 (0.10)</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td><strong>Interpersonal skills</strong>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IS-1. My cardiologist seemed supportive of my emotions.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>7 (6.6)</td>
<td>40 (37.7)</td>
<td>52 (49.1)</td>
<td>4.26 (0.10)</td>
<td>0.39</td>
<td>.76</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>3 (3.1)</td>
<td>6 (6.3)</td>
<td>27 (28.1)</td>
<td>58 (60.4)</td>
<td>4.42 (0.09)</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>4 (3.8)</td>
<td>0 (0.0)</td>
<td>6 (5.7)</td>
<td>37 (34.9)</td>
<td>59 (55.6)</td>
<td>4.38 (0.09)</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td><strong>IS-2. I was comfortable discussing my medical concerns.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>7 (6.6)</td>
<td>40 (37.7)</td>
<td>52 (49.1)</td>
<td>4.26 (0.10)</td>
<td>0.39</td>
<td>.54</td>
</tr>
<tr>
<td><strong>IS-3. My cardiologist displayed appropriate interpersonal skills.</strong></td>
<td>4 (3.8)</td>
<td>3 (2.8)</td>
<td>7 (6.6)</td>
<td>40 (37.7)</td>
<td>52 (49.1)</td>
<td>4.26 (0.10)</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td><strong>Supportive environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SE-1. My interaction with other in-office personnel was professional.</strong></td>
<td>5 (4.7)</td>
<td>0 (0.0)</td>
<td>10 (9.4)</td>
<td>43 (40.6)</td>
<td>48 (45.3)</td>
<td>4.22 (0.09)</td>
<td>0.32</td>
<td>.37</td>
</tr>
<tr>
<td>Tele</td>
<td>2 (2.1)</td>
<td>1 (1.0)</td>
<td>7 (7.3)</td>
<td>32 (33.3)</td>
<td>54 (56.3)</td>
<td>4.41 (0.09)</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>In-P</td>
<td>10 (9.4)</td>
<td>43 (40.6)</td>
<td>48 (45.3)</td>
<td>4.22 (0.09)</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td><strong>Average ratings of all items</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Respondents rated the in-person experience somewhat higher across all 11 individual items (Table 2 and Figure 1); the mean rating in telehealth for 8 of the 11 items was below the grand mean, while only 2 items (patient-centered communication–1 and interpersonal skills–3) were above the grand mean; by contrast, only 2 items (clinical competence–2 and interpersonal skills–1) showed a mean in-person rating below the grand mean (Figure 1). The paired difference in average response was significantly lower for telehealth ($z$=3.98, $P$<.001). Despite this trend, only the item relating to the perceived thoroughness of the clinical exam showed a significantly different pattern of responses between appointment types. However, there was no significant difference in mean response to the single item related to respondents’ overall perception of their telehealth or in-person experience ($z$=1.22, $P$=.22). These data are summarized in Table 2.
Figure 1. Changes in satisfaction ratings for individual survey items between telehealth and in-person. Points are expressed as the deviation of the mean of individual survey items from the grand mean of 4.30 for in-person (horizontal axis) and telehealth experiences (vertical axis). Labels reflect the survey domain and item number as indicated in Table 2. Points above and/or to the right of their respective axis indicate items whose mean rating was above the grand mean of all items, while those to the left and/or below indicate points with ratings below the grand mean. The lower right quadrant contains items for which in-person mean ratings were above the grand mean, while in telehealth were below the grand mean. CC: clinical competence; IS: interpersonal skills; PCC: patient-centered communication; SE: supportive environment.

All individual survey items showed significant positive correlations with respondents’ overall rating of their experience, across both telehealth and in-person formats, based on Bonferroni-adjusted criteria. For telehealth, Spearman correlations ranged from 0.49 for the item related to thoroughness of the clinical examination \((P<.001)\) to 0.32 for the item related to the interaction with in-office personnel \((P<.001)\). For the in-person experience, correlations ranged from 0.56 for the item relating to the clarity of instructions and treatment plans \((P<.001)\) to 0.40 for the item related to the appropriateness of the level of medical care provided \((P<.001)\). These data are summarized in Table 2.

Average ratings for all cardiologists across both telehealth and in-person formats was uniformly high; all means for both were above 4.0 on a five-point scale (Figure 2). In addition, all cardiologists showed minimal difference in mean ratings across the two appointment types (Figure 2). There was no significant difference among cardiologists in their patients’ perceptions of either their telehealth (KW statistic 6.24, \(df=6, P=.40\)) or in-person experience (KW statistic 3.75, \(df=6, P=.71\)). Similarly, there was no difference among cardiologists in the paired difference in telehealth versus in-person ratings (KW statistic 7.2, \(df=6, P=.30\)).
Perception: Advantages and Disadvantages to Telehealth

Reduced travel time was seen as a big advantage over traditional in-person appointments by 61 (57.5%) of the 106 respondents who participated in telehealth, and 94 (88.7%) viewed it at least somewhat of an advantage. Similarly, the majority (n=57, 53.8%) viewed reduced visit wait time as a big advantage, and 94 (88.7%) saw it as at least somewhat of an advantage. A similar percentage (n=87, 82.0%) saw travel cost savings as at least somewhat of an advantage to telehealth, including 43 (40.5%) who rated it as a big advantage. These data are summarized in Table 1. Respondents listed increased comfort, the ability to continue work, and lower risks of COVID-19 as additional benefits in the open response.

There was no relationship between communication modality (ie, phone, smartphone, computer, or tablet) and respondents’ overall rating of the telehealth experience ($\chi^2_8=6.14, P=.63$); similarly, there was no relationship between software platform and overall ratings ($\chi^2_2=0.91, P=.63$), though the majority of respondents (62.3%) indicated they did not remember the platform used. There was also no relationship between respondents’ travel time to in-person appointments and their overall rating of the telehealth experience (Spearman $r=0.02, df=1, P=.24$). Of 120 respondents, 100 (83.0%) indicated they would at least consider using telehealth going forward.

Among the 106 respondents who participated in telehealth, fewer than 10% (range 4-9 respondents, 3.8%-8.5%) rated any of the potential issues as a big disadvantage; by contrast, individual survey items were rated as not a disadvantage by 67%-86% (range 71-91) of respondents, based on their experience. Privacy concerns were seen as the least problematic of the potential issues, with only 15 (14.2%) respondents reporting this as at least somewhat of a disadvantage. Poor internet connectivity was of most concern, rated as at least somewhat of a factor by 35 (33.0%) respondents. These data are summarized in Table 1. Responses collected via open response included a lack of hands-on attention, difficulty communicating, and a lack of end-of-visit paper summaries as additional disadvantages.

### Discussion

Access to Telehealth Offers Both Opportunities and Challenges

This study takes advantage of the natural experiment provided by the COVID-19 pandemic to explore the utility of telehealth from the patient perspective. We found both opportunities and challenges related to accessibility, and the modality is perceived by patients as a viable alternative to in-person office visits and patients saw clear benefits to its use. Our results have implications for cardiology practices moving forward but should be interpreted with caution due to sampling constraints and the unique context of the global pandemic.

Internet and technology access do not seem to be significant barriers to the use of telehealth. Of the 193 initial respondents, 55 (28.4%) reported declining to use telehealth. However, among the 32 respondents who declined and reported factors, only a small percentage (n=2, 6.2%) cited access to technology as a factor in their decision. Of the 106 respondents who participated in telehealth, a similarly low percentage (n=8, 7.5%) viewed internet connectivity as a big disadvantage, though a more substantial 25.5% (n=27) did cite it as somewhat of a disadvantage. Nevertheless, patients expressed a fairly high level of satisfaction with telehealth, in terms of both average ratings among items and overall rating of their experience. Similarly, more than 70% of respondents reported unfamiliarity with technology (both hardware and software) as not being a
factor in declining telehealth or as a disadvantage by those who participated (n=82, 77.4% and n=76, 71.7%, respectively). These findings suggest that, even during a period of rapid and unplanned change, internet access and use of technology are likely manageable issues for most patients and that continued, intentional efforts on the part of governments, health care systems, and corporate providers to address access disparities will only improve the situation moving forward.

However, there is some evidence that it may be harder to coordinate telehealth appointments, at least initially. Although respondents did not indicate significant issues in navigating or communicating as part of their telehealth appointments, our data do suggest there was some ambiguity about the need for or opportunity to participate in telehealth. Of the 32 respondents who did not participate in telehealth, 27 (84.4%) cited not having an appointment as at least somewhat of a factor in their decision. However, all patients invited to participate in the study had an appointment scheduled with their cardiologist prior to the COVID-19-related executive orders prohibiting in-person delivery of nonacute health care services; these appointments were shifted to a telehealth format. The most common reason patients did not meet their telehealth appointment was inability of the WKHL office to contact patients the day of their appointment, and we suspect miscommunication between the WKHL office and the patients or patients’ family members regarding changes in the appointment modality as the possible reason for this. Going forward, it will be important for providers to ensure consistent and reliable communication with patients to minimize any confusion regarding appointments.

Telehealth Is Perceived as a Viable Alternative to In-Person Cardiology Appointments

Patterns of Satisfaction Are Consistent Across Modalities

There was no significant shift in rankings of patient satisfaction scores between modalities. Although satisfaction scores decreased somewhat in telehealth for all items, the decreases were generally modest and consistent. This suggests that the different modalities do not present qualitatively different challenges to establishing a physician-patient relationship, though more intentional effort may need to be applied across the board to ensure that patients perceive telehealth as offering an equivalent standard of care.

Physicians Seem to Be Able to Adapt Well

Satisfaction scores were high and consistent among all 7 cardiologists represented in the sample. Despite having little or no previous experience with telehealth, all physicians appeared to operate effectively within the new environment. On a broader scale, there were few if any differences in patient satisfaction scores among the four survey domains of the physician-patient experience, both within and among telehealth and in-person modalities.

The Clinical Exam Is an Issue That Needs to Be Addressed

The only item that showed a significant decrease in patient satisfaction between in-person and telehealth visits was the perceived thoroughness of the clinical exam. Our patient population included a substantial number of older and rural individuals, many with limited technology abilities, limited access to technology, and limited access to broadband connection. This translated into a significant proportion of telehealth visits done without face-to-face evaluation, which might have contributed to a lower scoring on the physical examination component.

This finding is also consistent with existing concerns regarding telehealth in specialty fields [5]. It is clear that, if use of telehealth is to expand within cardiology or other similar fields, multiple mechanisms must be put in place to enable physicians to collect necessary clinical data remotely. Such remote patient monitoring solutions might include remote clinical stations located in partner clinics nearer to patients’ homes or use of smartphone apps that record heart rate, blood pressure, pulse oximetry, or electrocardiogram data and delivering those wirelessly to the physician [28].

Patients See Clear Advantages to Using Telehealth

More than 80% of the 106 respondents identified time (n=94, 88.7%) and cost savings (n=87, 82.1%) as either somewhat or a primary advantage of telehealth, and overall satisfaction with telehealth was independent of the distance traveled by respondents to in-person appointments. This suggests that the perceived time and cost savings are threshold benefits that positively impact the majority of patients more or less equally. By contrast, privacy concerns were not viewed as a factor either by those who participated in telehealth or those who opted out. This pattern suggests that time and cost efficiency for patients should be a primary concern when implementing telehealth and that sensitive issues such as privacy protection can be readily accommodated.

Limitations

Our study has some unavoidable limitations, due to its natural experiment dimension and the desire for real-time rapid response. The reliance on online delivery of the survey may well have limited our response rate, especially among those individuals less comfortable with or having limited access to technology. However, the ability to generate data on patient satisfaction in real time, as a means of rapidly assessing the mandated shift to telehealth, justifies its use. In any case, we appear to have captured a representative sample of our patient population, both demographically and in terms of accessibility (ie, telehealth no-shows), making the trends and relationships in our data worthy of further consideration. Moreover, post hoc power analyses indicated that our sample sizes were sufficient to establish a level of statistical power >0.95 for comparisons between telehealth and in-person visits.

Although our data highlight relevant lessons for the continued or expanded use of telehealth in cardiology, we must also be cautious. Satisfaction ratings of in-person appointments may be less reliable (and perhaps inflated) due to differences in reporting period; that is, we asked respondents to rate in-person experiences that occurred less recently than telehealth experiences. Longer reporting periods cause respondents’ ratings to be more affected by the most intense or recent experiences, while the impact of milder experiences is attenuated [29,30].
On the other hand, the dangers of COVID-19, especially for these patients who are at risk, nearly ensures a positive bias toward telehealth, which may disappear somewhat or entirely if and when the health care delivery system returns to more "normal" operation. As a result, we may have observed less difference between satisfaction with telehealth and in-person appointments than we might have originally expected (or might expect to see in the future). Once the fear related to the COVID-19 pandemic subsides, will patients still feel as positive about their experiences with telehealth?

These caveats suggest that, although we could expect the patterns among individual survey items to hold, we should be cautious in assuming that the degree of equivalency observed between telehealth and in-person satisfaction can be generalized to new health care delivery contexts. They also argue for considering even nonsignificant trends, as these may be indicative of differences that could become accentuated in a more normal environment. Finally, they highlight the need for randomized controlled trials to truly evaluate differences between in-person and telehealth experiences.

**Conclusions: Future Application of a Telehealth Model in Specialty Fields Such as Cardiology**

The overall level of satisfaction expressed with telehealth and perceived time- and cost-saving benefits identified by patient indicate that it can play an increasing role in providing health care access and services beyond COVID-19, particularly in rural areas. As such, the efficacy of telehealth needs to be better examined, especially in medical specialty fields, and patient and provider perception of telehealth needs to be evaluated to determine if it is worth expanding into regular practice. Increased literature on telehealth use in rural populations will hopefully aid in determining the best course of action in addressing health care disparities in a substantial part of the United States.

**Acknowledgments**

The authors would like to acknowledge the efforts of Dr Muhammad Akbar, Dr Jacqueline Dawson, Dr Mohammed Kazimuddin, Nick Dylan, and Dr Patrick Jones for their help with literature review and developing the survey, and John Michael Mills in preparation of this manuscript.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

EMR: electronic medical record
KW: Kruskal-Wallis
PACT: Patient Assessment of Communication During Telemedicine
WKHL: Western Kentucky Heart and Lung

Please cite as:
Patient Perspectives With Telehealth Visits in Cardiology During COVID-19: Online Patient Survey Study
JMIR Cardio 2021;5(1):e25074
URL: http://cardio.jmir.org/2021/1/e25074/
doi: 10.2196/25074
PMID:33385234

Edited by G Eysenbach; submitted 16.10.20; peer-reviewed by S Agnihotri, IV George; comments to author 03.11.20; revised version received 23.11.20; accepted 24.12.20; published 22.01.21.
Predicting Cardiovascular Risk Using Social Media Data: Performance Evaluation of Machine-Learning Models

Anietie U Andy¹, PhD; Sharath C Guntuku¹,², PhD; Srinath Adusumalli³,⁴, MD, MSc; David A Asch³,⁵,⁶, MD, MBA; Peter W Groeneveld⁵,⁶, MD, MS; Lyle H Ungar⁴, PhD; Raina M Merchant¹,³,⁷, MD, MSHP

¹Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA, United States
²Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States
³Penn Medicine Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States
⁴Division of Cardiovascular Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
⁵Center for Health Equity Research and Promotion, Corporal Michael J Crescenz VA Medical Center, Philadelphia, PA, United States
⁶Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
⁷Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Anietie U Andy, PhD
Penn Medicine Center for Digital Health
University of Pennsylvania
Philadelphia, PA
United States
Phone: 1 202 486 4095
Email: Anietie.Andy@pennmedicine.upenn.edu

Abstract

Background: Current atherosclerotic cardiovascular disease (ASCVD) predictive models have limitations; thus, efforts are underway to improve the discriminatory power of ASCVD models.

Objective: We sought to evaluate the discriminatory power of social media posts to predict the 10-year risk for ASCVD as compared to that of pooled cohort risk equations (PCEs).

Methods: We consented patients receiving care in an urban academic emergency department to share access to their Facebook posts and electronic medical records (EMRs). We retrieved Facebook status updates up to 5 years prior to study enrollment for all consenting patients. We identified patients (N=181) without a prior history of coronary heart disease, an ASCVD score in their EMR, and more than 200 words in their Facebook posts. Using Facebook posts from these patients, we applied a machine-learning model to predict 10-year ASCVD risk scores. Using a machine-learning model and a psycholinguistic dictionary, Linguistic Inquiry and Word Count, we evaluated if language from posts alone could predict differences in risk scores and the association of certain words with risk categories, respectively.

Results: The machine-learning model predicted the 10-year ASCVD risk scores for the categories <5%, 5%-7.4%, 7.5%-9.9%, and ≥10% with area under the curve (AUC) values of 0.78, 0.57, 0.72, and 0.61, respectively. The machine-learning model distinguished between low risk (<10%) and high risk (>10%) with an AUC of 0.69. Additionally, the machine-learning model predicted the ASCVD risk score with Pearson \( r = 0.26 \). Using Linguistic Inquiry and Word Count, patients with higher ASCVD scores were more likely to use words associated with sadness \( r = 0.32 \).

Conclusions: Language used on social media can provide insights about an individual’s ASCVD risk and inform approaches to risk modification.

(JMIR Cardio 2021;5(1):e24473) doi:10.2196/24473

KEYWORDS
ASCVD; machine learning; natural language processing; atherosclerotic; cardiovascular disease; social media language; social media
**Introduction**

Secondary prevention approaches have improved the longevity of patients with cardiovascular (CV) disease; however, risk factors and adverse health behaviors (e.g., physical inactivity, smoking) are highly prevalent, and <1% of adults in most contemporary series meet all factors of ideal CV health [1]. The logistics and practicalities of meeting the goal of ideal CV health have not been clearly elucidated. Practice guidelines recommend using atherosclerotic cardiovascular disease (ASCVD) pooled cohort equations (PCESs) [2] or other prediction tools to classify patients’ risk of CV disease and the need for risk-reducing therapies such as statin medications [3]. There is also an increasing focus on identifying markers that provide better measures of risk. To best prevent ASCVD, it is important to precisely determine an individual’s 10-year risk for ASCVD. As digital platforms are increasingly used to document health behaviors, data from digital sources may provide a window into manifestations of novel risk factors and can provide complementary data to characterize existing risk factors.

Social media data in the form of posts, photos, and “likes” can provide information about individuals’ daily activities and behaviors. Social media has been used to track heart disease mortality rates [4] and depression [5]. Data on social media platforms are generated at a fast rate. Accessing these data from consenting individuals offers an opportunity to collect and analyze these data in real time. This information could facilitate identification of earlier signals of disease development or exacerbation, and timely tracking of the health of individuals and the collective health of a community [4-9]. The data are generally unscripted and spontaneous, and can therefore provide information that is different from standard survey assessments. Another potential use of data from digital platforms is that they can be used for direct intervention, so that the same platforms that are being used to assess insights can also be used to deliver targeted health information or evaluate information delivery.

The potential of social media data for CV health lies in tracking, codifying, and better understanding the hard-to-measure lifestyle choices, along with exposures related to diet, exercise, smoking, and other factors that can significantly contribute to the development and progression of heart disease. At present, measuring many of these behaviors is dependent on self-report and recall [4]. Yet, posts or images from digital media could better inform a patient-provider discussion about how to change actual dietary choices and consumption. Incorporating data from digital sources has the potential to enhance our approach for characterizing individuals’ risk and tailoring management, as a new type of precision medicine.

We sought to use social media data from consenting individuals to predict ASCVD risk reported in an electronic medical record (EMR), and to characterize differences in posts relative to four categories based on the 10-year primary risk from ASCVD risk scores.

**Methods**

**Data and Design**

This was a retrospective analysis of social media and EMR data of consenting patients. This study was approved by the University of Pennsylvania Institutional Review Board.

Recruitment for compiling the social medimome dataset began in March 2014 and included patients from inpatient and outpatient settings across two urban academic medical centers. Participants in this dataset consented to sharing access for selecting historical data from their social media accounts (e.g., Facebook, Twitter, Instagram) and access to their medical record data. Data are stored in a Health Insurance Portability and Accountability Act (HIPAA)-compliant secure server and participants can elect to discontinue sharing data at any time point. Additional information about this dataset is published elsewhere [10,11].

Following recommendations of Goff et al [12], which outlines the process of developing a risk equation for predicting the 10-year ASCVD risk of individuals between the ages of 40 and 79 years, from our dataset, we identified patients aged 40 to 79 years and without a prior history of ASCVD documented in their EMR. Of these patients, we identified 181 with a calculated 10-year primary risk of ASCVD score in their EMR. For all these patients, demographics (age, race, gender) were also extracted from the EMR along with ASCVD scores. We retrieved Facebook status updates up to 5 years prior to enrollment for all users.

Table 1 shows the demographic information of patients included in our analysis. Of the 181 participants meeting the criteria for this study, the majority were women and Black, and the average age was 50 years. The participants had 159,958 Facebook posts overall (mean 884, SD 3227).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>104 (57.5)</td>
</tr>
<tr>
<td>White</td>
<td>64 (35.4)</td>
</tr>
<tr>
<td>Other</td>
<td>13 (7.2)</td>
</tr>
<tr>
<td><strong>Men, n (%)</strong></td>
<td>48 (26.5)</td>
</tr>
</tbody>
</table>

We used two approaches to process language from social media posts for inclusion in a regression model. Specifically, language features from posts were derived using (a) open vocabulary topics and (b) dictionary-based psycholinguistic features. These derived language features were then used to predict the patients’
10-year ASCVD risk scores and to distinguish patients with different ASCVD risk scores.

**Open Vocabulary Approach**

The open vocabulary approach uses latent Dirichlet allocation (LDA) [13], which is a natural language-processing method that is used to analyze the co-occurrences of words in text (in this case Facebook posts). Distinct groupings of these words represent topics (eg, groups of co-occurring words) and these topics can be labeled based on their content. For example, the model could cluster the words “dinner,” “cheese,” “eat,” “made,” and “food” as a reference to food by utilizing the similarities in the distributional properties in the Facebook posts.

We generated 20 topics using Facebook posts from all of the users in our dataset. Each user was represented as a 20-dimensional vector based on the probability of each topic in all users’ posts. Multimedia Appendix 1 shows the LDA topics and 10 words associated with each topic. To determine the number of topics, consistent with prior work [14,15], we varied the number of topics using 10, 20, 50, 75, and 100 topics, respectively; 20 topics had the most coherent topic themes when reviewed by one of the coauthors.

**Dictionary-Based Approach**

The dictionary-based psycholinguistic approach uses language from Facebook posts to identify the prevalence of predefined word categories represented in the Linguistic Inquiry and Word Count (LIWC) dictionary [16]. LIWC represents a dictionary of 73 different psycholinguistic word categories such as topical categories and emotions. For each user, the rate of words that occurred in a given LIWC category was measured and included as input in a model to predict ASCVD risk as described below.

**Predicting ASCVD Risk Scores Using Social Media Language**

We sought to investigate the discriminatory power of predicting a patient’s 10-year ASCVD risk using language features derived from Facebook posts. We extracted the features described using an open-vocabulary approach and trained a logistic regression model, as implemented in Python 3.4 scikit-learn [17], to predict ASCVD risk scores using 5-fold cross-validation. We defined the outcome in three different ways.

**Model 1**

The multiclass logistic regression model on Facebook posts was trained to classify patients in four different categories (<5%, 5%-7.4%, 7.5%-9.9%, ≥10%) based on their ASCVD risk scores. The model was able to delineate patients in the lowest risk category (<5%) from patients in other categories with an AUC of 0.78. The model delineated patients in the categories 5%-7.4%, 7.5%-9.9%, and ≥10% from those in other categories, as shown in Table 2.

<table>
<thead>
<tr>
<th>Category</th>
<th>AUC (language only)</th>
<th>AUC (age only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5%</td>
<td>0.52</td>
<td>0.78</td>
</tr>
<tr>
<td>5%-7.4%</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>7.5%-9.9%</td>
<td>0.45</td>
<td>0.72</td>
</tr>
<tr>
<td>≥10%</td>
<td>0.59</td>
<td>0.61</td>
</tr>
</tbody>
</table>

**Model 2**

Using the linear regression model on Facebook posts, we predicted the ASCVD risk score of patients with \( r=0.26 \) (\( P<.001 \)).

**Model 3**

The logistic regression model delineated patients with a high risk (≥10%) of ASCVD from those with a low risk (<10%) with an AUC of 0.69.

In 2013, the American Heart Association and American College of Cardiology put forth the ASCVD PCEs [2], which can be used to predict an individual’s 10-year risk of ASCVD. Therefore, in Model 1, ASCVD risk was set as a categorical variable. We categorized patients into the following different thresholds: <5%, 5%-7.4%, 7.5%-9.9%, and ≥10%. We trained a multiclass logistic predictive model to predict these four categories of ASCVD risk scores. The prediction performance is reported as the area under the receiver operating characteristic curve (AUC).

For Model 2, the ASCVD risk score of patients was applied as a continuous variable rather than as a categorical variable that was used in Model 1. The performance of Model 2 was assessed using the Pearson correlation coefficient (r).

Identifying patients with high risk (≥10%) of ASCVD is of interest to clinicians. Therefore, in Model 3, we treated ASCVD risk as a dichotomous variable and built a logistic regression model to distinguish the high-risk category using language compared to low ASCVD scores (ie, <10%). Additionally, we used LIWC to distinguish the different features associated with high-risk patients by correlating the LIWC category feature of patients from their social media posts and whether they are in the high-risk (>10%) or low-risk (<10%) categories; we measured the effect size using Cohen d. To indicate significant correlations, we used Benjamini-Hochberg \( P \) value correction with a significance threshold of \( P<.001 \).

**Results**

**Predicting ASCVD Risk Score Using Social Media Language**

**Model 1**

The multiclass logistic regression model on Facebook posts was trained to classify patients in four different categories (<5%, 5%-7.4%, 7.5%-9.9%, ≥10%) based on their ASCVD risk scores. The model was able to delineate patients in the lowest risk category (<5%) from patients in other categories with an AUC of 0.78. The model delineated patients in the categories 5%-7.4%, 7.5%-9.9%, and ≥10% from those in other categories, as shown in Table 2.
Identifying Differentially Expressed Language Features According to High and Low ASCVD Scores

The sadness LIWC category was most strongly associated with the high ASCVD risk category (≥10%) at a Benjamini-Hochberg–corrected significance level of \( P \leq 0.001 \) and Pearson \( r = 0.32 \). None of the other LDA topics or LIWC categories was significantly associated with high and low ASCVD risk.

Discussion

Principal Findings

Language from Facebook posts has the potential to distinguish patients based on their calculated 10-year ASCVD risk score categorization and actual risk score. Although social media data are unlikely to replace traditional approaches for predicting CV risk, these findings suggest that such data can potentially provide supplemental information about an individual’s lifestyle and behavior, which can complement our understanding of contributors to long-term CV risk. More than 2 billion people share information about their daily lives on social media platforms, which can include information about what they eat and drink, if they smoke, when they exercise, what their lab results are, and other factors associated with Life’s Simple 7 [18]. However, less is known about how much of this information is noise or if there is an actual relevant signal in the volumes of data in online chatter such as Facebook, where individuals often reveal information about themselves. Additionally, prior work has demonstrated that social media data can be used to predict several medical conditions such as diabetes and mental health conditions [4,5].

The potential opportunity in exploring social media data is that this emerging data source could include data about behavior and lifestyle that might not have been reported to clinicians. There is still a gap in how this would be implemented in clinical practice, and would require further evaluation of feasibility, acceptability, and interpretability. These data are unlikely to replace the existing risk score input but rather may provide complementary adjunct data. Prior work has explored the contribution of nonclinical factors (eg, patient interviews about socioeconomic status, health status, adherence, psychosocial characteristics) in predicting CV outcomes (eg, congestive heart failure readmissions). The model performance overall was poor, although patient-reported information extended the predicted ranges of rates of readmission and slightly improved model discrimination [19]. Social media data in the form of photos, videos, and likes [20,21] have been used to predict users’ personality [22], mental health, and other behaviors. Consequently, future work could use multiple modalities of user-generated content to model the ASCVD risk score.

In our patient cohort, a high ASCVD risk score was associated with increased use of “sad” language on Facebook. This is consistent with research demonstrating that depression is more prevalent in populations with CV disease, and is predictive of adverse outcomes (such as myocardial infarction and death) among populations with preexisting CV disease [23].

In our analysis, the AUC for Model 1 indicated low accuracy. A potential reason for this is that we used data from individuals between the ages of 40 and 79 years, and individuals in this age group do not post as much on social media compared to younger individuals. Accordingly, in our dataset, some users had fewer posts, leading to low accuracy from the AUC. We hypothesize that with more posts (ie, more words), our models will perform better.

We compared Models 1 and 3 together to determine which performed better at predicting the ASCVD risk score of individuals. Toward this end, we computed the micro AUC of Model 1 and compared it to that of Model 3, which was 0.66 and 0.69, respectively. This suggests that Model 3 is more reliable at predicting ASCVD risk compared to Model 1.

The findings of this study offer promise for using emerging digital data sources for identifying risk factors. This moves beyond what is simply reported by patients to what may be revealed when looking at a diary of information over multiple time points. This could aid clinicians in providing individualized recommendations for managing risk factors that contribute to heart disease.

Limitations

This study has several limitations. The study cohort was primarily female and African American. Our analysis used posts from patients with at least 200 words in their Facebook posts, and therefore we cannot extrapolate about those who used social media less or did not consent to share; we used 200 words because prior work on using social media for predicting individuals’ traits determined that for good and stable predictive performance when working with social media data, data from users with 200 words or more on Facebook should be used [24,25]. Our sample was also limited to those with an ASCVD risk score in a single health system EMR, and therefore we may have missed individuals with a risk score in another EMR or that may not have had a risk score calculated in our EMR.

Conclusion

We show that language from Facebook posts can be used to predict an individual’s 10-year risk for ASCVD. Specific information in posts could help to guide clinicians in better understanding lifestyles and behaviors, and in counseling patients about heart disease risk.

Acknowledgments

This study is funded by the National Heart Lung and Blood Institute (NHLBI) of the National Institutes of Health (NIH) (1R01HL141844-01A1) to RM (principal investigator), DA, LA, and PG (coinvestigators).
Conflicts of Interest

DA is a partner and part owner of VAL Health, and is a US government employee. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Twenty topics generated from our dataset.

[DOCX File, 14 KB - cardio_v5i1e24473_app1.docx]

References


Abbreviations

ASCVD: atherosclerotic cardiovascular disease  
AUC: area under the receiver operating characteristic curve  
CV: cardiovascular  
EMR: electronic medical record  
LDA: latent Dirichlet allocation  
LIWC: Linguistic Inquiry and Word Count  
PCE: pooled cohort equations

©Anietie U Andy, Sharath C Guntuku, Srinath Adusumalli, David A Asch, Peter W Groeneveld, Lyle H Ungar, Raina M Merchant. Originally published in JMIR Cardio (http://cardio.jmir.org), 19.02.2021. 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 http://cardio.jmir.org, as well as this copyright and license information must be included.
Heart Rate Measurements in Patients with Obstructive Sleep Apnea and Atrial Fibrillation: Prospective Pilot Study Assessing Apple Watch’s Agreement With Telemetry Data

Pauline Huynh1*, BA; Rongzi Shan2,3*, BSc; Ngozi Osuji2, MD, MPH; Jie Ding2, PhD; Nino Isakadze1,2, MD; Françoise A Marvel1,2, MD; Garima Sharma1,2, MD; Seth S Martin1,2, MD, MHS

1School of Medicine, Johns Hopkins University, Baltimore, MD, United States
2Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States
3David Geffen School of Medicine, University of California, Los Angeles, CA, United States

*these authors contributed equally

Corresponding Author:
Seth S Martin, MD, MHS
School of Medicine
Johns Hopkins University
600 N Wolfe St, Carnegie 591
Baltimore, MD, 21287
United States
Phone: 1 410 502 0469
Email: smart100@jhmi.edu

Abstract

**Background:** Patients with obstructive sleep apnea (OSA) are at a higher risk for atrial fibrillation (AF). Consumer wearable heart rate (HR) sensors may be a means for passive HR monitoring in patients with AF.

**Objective:** The aim of this study was to assess the Apple Watch’s agreement with telemetry in measuring HR in patients with OSA in AF.

**Methods:** Patients with OSA in AF were prospectively recruited prior to cardioversion/ablation procedures. HR was sampled every 10 seconds for 60 seconds using telemetry and an Apple Watch concomitantly. Agreement of Apple Watch with telemetry, which is the current gold-standard device for measuring HR, was assessed using mixed effects limits agreement and Lin’s concordance correlation coefficient.

**Results:** A total of 20 patients (mean 66 [SD 6.5] years, 85% [n=17] male) participated in this study, yielding 134 HR observations per device. Modified Bland–Altman plot revealed that the variability of the paired difference of the Apple Watch compared with telemetry increased as the magnitude of HR measurements increased. The Apple Watch produced regression-based 95% limits of agreement of 27.8 – 0.3 × average HR – 15.0 to 27.8 – 0.3 × average HR + 15.0 beats per minute (bpm) with a mean bias of 27.8 – 0.33 × average HR bpm. Lin’s concordance correlation coefficient was 0.88 (95% CI 0.85-0.91), suggesting acceptable agreement between the Apple Watch and telemetry.

**Conclusions:** In patients with OSA in AF, the Apple Watch provided acceptable agreement with HR measurements by telemetry. Further studies with larger sample populations and wider range of HR are needed to confirm these findings.

(JMIR Cardio 2021;5(1):e18050) doi:10.2196/18050

KEYWORDS

mHealth; wearables; atrial fibrillation; obstructive sleep apnea; digital health

Introduction

Atrial fibrillation (AF) is the most common clinically significant cardiac arrhythmia, with a lifetime risk of 1 in 4 among individuals over the age of 40 and about 1 in 3 among individuals over the age of 55, thereby posing substantial public health and economic burden [1,2]. AF is associated with significant cardiovascular and cerebrovascular morbidity and
mortality, including a fivefold risk of thromboembolic complications such as stroke [1,3]. Because of its episodic, paroxysmal, and minimally symptomatic nature, the diagnosis of AF is often delayed, with nearly 1 in 5 diagnoses occurring at the onset of acute stroke [4]. It is estimated that nearly 700,000 people in the United States alone have undiagnosed AF due to its “clinically silent” nature, presenting a diagnostic challenge for clinicians [5].

Of particularly high risk for developing AF are individuals with sleep breathing disorders, including obstructive sleep apnea (OSA). A strong association between OSA and AF has been consistently observed in both epidemiological and clinical cohorts, with patients with OSA being 2 to 4 times more likely to develop AF compared to those without OSA [6-8]. Gami et al [9] reported significantly higher prevalence (49% vs 32%) of OSA and a strong association (adjusted odds ratio of 2.19) between OSA and AF in patients undergoing electrical cardioversion as compared to patients without AF. Moreover, a comorbid diagnosis of OSA is predictive of AF recurrences after catheter ablation or electrical cardioversion of AF [7,10].

Recently, the growing prevalence and adoption of digital health tools, including mobile devices with physiologic sensors (eg, “wearables”), have caught the attention of industry giants in the technology sector and clinicians who see opportunities for synergy in subclinical AF detection. This is evidenced by the rapid development and release of wearables for AF detection, including the Apple Watch Series 4 (Apple Inc.), KardiaBand and KardiaMobile (AliveCor), Hexoskin (Carré Technologies Inc.), and QardioCore (Qardio Inc.) [11]. Of these, only the Apple Watch Series 4 and KardiaMobile have been FDA cleared for AF detection [12,13], although many still list claims promoting heart health and wellness. Furthermore, ownership of wearables has more than doubled between 2014 and 2018 (from 25.1 million to 51.9 million users), and is further projected to increase with nearly half of the American public showing interest in future ownership [11,14].

Many wearables monitor heart rate (HR) through an optic technology known as photoplethysmography (PPG), in which sensors detect and measure pulsatile light absorption in the vasculature beneath the skin as a proxy for the cardiac cycle [15]. While this intersection in health technology has spurred numerous validation studies in the detection of AF [3,14,16], little is known about the accuracy of PPG technology in measuring HR during AF. Preliminary work by a single group in Australia suggests that during AF episodes, smart watches underestimate HR over 100 beats per minute (bpm) when compared to electrocardiogram (ECG) or Holter monitoring [17,18]. Similarly, as wearables evolve to accurately detect AF and bring users into the health care system, little research exists on how these technologies may also be used to help patients assess their AF management plans, which may include a rate control strategy and detection of rapid ventricular response (RVR).

In this pilot study, we assessed the Apple Watch’s agreement with telemetry as the gold standard in measuring HR in patients with OSA in AF. We chose to recruit patients with OSA given their higher likelihood of having a co-diagnosis of AF [19] and because we had encountered in clinical practice patients with OSA who had self-identified AF with RVR by a fast HR on their Apple Watch. We hypothesized that the Apple Watch would measure HR accurately when compared to standard ECG monitoring in patients with OSA in AF.

Methods

Study Approval

This study was approved by the Johns Hopkins Medicine Institutional Review Board. Apple Inc. was not involved in the design, implementation, data analysis, or manuscript preparation of the study.

Study Design

In this prospective pilot study, patients aged 18 and older with OSA in AF episodes confirmed on ECG were identified via electronic health record screening and prospectively recruited prior to cardioversion and AF ablation procedures at Johns Hopkins Hospital between November 2018 and May 2019. Diagnosis of OSA was determined by chart review, and patients with objective clinical documentation of (1) current continuous positive airway pressure (CPAP) device use, (2) polysomnogram results showing OSA, or (3) both were considered eligible. Patients were excluded if they had implantable pacemakers, defibrillators, loop recorders, heart block, or tachycardia not attributable to AF. In addition, patients who were hemodynamically unstable or under contact precautions for infection control were excluded.

Data Collection

Eligible patients were approached prior to their procedures and provided informed written consent. AF was confirmed by a 12-lead ECG performed minutes prior to HR data collection. Participants wore a first-generation Apple Watch (model A1554), which was provided by the study team for the duration of the study. The Apple Watch face and telemetry monitor (CARESCAPE Monitor B650; GE Healthcare) were observed concomitantly under video recording in the presence of a study co-investigator (RS) for 90 seconds. After excluding the first 30 seconds of data to allow time for the watch’s HR monitor to equilibrate, HR measurements were sampled every 10 seconds for 60 seconds, yielding a total of 7 observations per participant per device (Apple Watch and telemetry). In addition, we documented the following relevant clinical data: cardiac history, cardiovascular medications, OSA treatment, nature of AF diagnosis, and demographic characteristics using the electronic health record. Full study flow can be found in Figure 1.

http://cardio.jmir.org/2021/1/e18050/
Statistical Analysis

Descriptive statistics were performed for the baseline characteristics, using frequencies (percentages) to describe categorical variables and mean (SD) or median (interquartile range) to describe continuous variables. Using the telemetry-determined HR as the gold standard, the Apple Watch was assessed for accuracy by calculating the paired difference between the measures. We first checked the mean constant bias assumption by visualizing the modified Bland–Altman plot accounting for repeated measures per patient (Figure 2). The mean bias appeared to be greater for higher HR measurements than for lower ones and log transformation of the data did not remove such relationship. We then analyzed the paired differences of Apple Watch compared with telemetry using a mixed effects regression model, with patients as a random effect and the averaged HR as the fixed effect. The paired difference was modeled in the following form [17]:

\[ \text{Diff}_{ijk} = \alpha + r_i + \beta_kA + e_{ij} \]

\[ ri \sim N(0, \sigma_i^2), e_{ij} \sim N(0, \sigma_e^2) \]

where \( \text{Diff}_{ijk} \) represents the \( j \)th paired difference in HR between devices in patient \( i \) given \( k \) value of the true (average) measurement; \( \alpha \) is the constant intercept; \( r_i \) is the random effect of the \( i \)th patient; \( \beta_k \) is the fixed effect of average of 2 measurements: and \( e_{ij} \) is the error for paired difference \( j \) on patient \( i \). The regression of \( \text{Diff}_{ijk} \) on the fixed effect of average of measurements gave the following:

\[ \text{Diff}_{ijk} = 27.7922 - 0.3332A \]

The coefficient of \(-0.3332\) was statistically significant (\( P<.05 \)) and further confirmed the average difference was related to the magnitude. We thus calculated the regression-based 95% limits of agreement as 27.7922 – 0.3332A – 1.96 × SD (of the residuals; lower limit) and 27.7922 – 0.3332A + 1.96 × SD (of the residuals; upper limit). An estimate of SD (7.6407) was calculated by the square root of total variance for all observations including the estimated between-patient variance and within-patient variance. Data were analyzed using the \textit{nlme} package of R software version 3.6.1 (R Foundation).
Results

Over the course of 6 months, we screened 201 consecutive patients who were scheduled for cardioversion and AF ablation procedures. Of these patients, 35 met full eligibility criteria and 22 patients were enrolled into the study (Figure 1). Demographic and clinical characteristics of the study participants are shown in Table 1. The mean age was 66 (SD 6.5) years and 85% (n=17) were male. Among the 20 participants analyzed, 3 (15%) had RVR. The majority of participants had persistent AF (14/20, 70%) and were prescribed antiarrhythmic (11/20, 55%), rate control (15/20, 75%), or anticoagulant (19/20, 95%) medications at the time of study enrollment.

Of the 280 possible HR measurements, 268 were recorded (95.7%). The first participant had 4 out of 14 recordings because the protocol was subsequently changed to capture a greater number of time points over 60 seconds of monitoring. A subsequent participant had 12 out of 14 recordings due to a failure to capture the entire 60 seconds of continuous monitoring on video. HR recordings ranged from 49 to 146 bpm from telemetry and 55 to 127 bpm from the Apple Watch. Figure 3 shows the standard deviation of the difference in paired measurement for each patient against the average measurement for that patient. As mentioned in the “Statistical Analysis” section, there was a suggestion that the variability of the difference increased as the magnitude of HR measurements increased. After performing the mixed effects regression model, we found that the 95% limits of agreement were calculated as $27.7922 - 0.3332A - 1.96 \times 7.6407$ (lower limit) and $27.7922 - 0.3332A + 1.96 \times 7.6407$ (upper limit), where $A$ is the magnitude (average of 2 methods) of HR. Based on this approach, the fit was greatly improved, particularly for higher HR. The Apple Watch had 95% of differences fall within 15.0 bpm above and 15.0 bpm below telemetry measurements. Lin’s concordance correlation coefficient between the Apple Watch and telemetry is 0.88 (95% CI 0.85-0.91).
Table 1. Participant characteristics (N=20).

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>66.0 (6.5)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>33.2 (4.8)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17 (85)</td>
</tr>
<tr>
<td>Female</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>16 (80)</td>
</tr>
<tr>
<td>Black</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Atrial fibrillation, n (%)</td>
<td></td>
</tr>
<tr>
<td>Paroxysmal</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Persistent</td>
<td>14 (70)</td>
</tr>
<tr>
<td>CHAD-VASC(^a) score, n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8 (40)</td>
</tr>
<tr>
<td>2</td>
<td>5 (25)</td>
</tr>
<tr>
<td>3</td>
<td>7 (35)</td>
</tr>
<tr>
<td>Antiarrhythmic medications, n (%)</td>
<td></td>
</tr>
<tr>
<td>Amiodarone</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Dofetilide</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Sotalol</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Propafenone</td>
<td>1 (5)</td>
</tr>
<tr>
<td>None</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Anticoagulant medications, n (%)</td>
<td></td>
</tr>
<tr>
<td>Rivaroxaban</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Apixaban</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Dabigatran</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Warfarin</td>
<td>3 (15)</td>
</tr>
<tr>
<td>None</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Rate control medications, n (%)</td>
<td></td>
</tr>
<tr>
<td>Nadolol</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Metoprolol succinate</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Metoprolol tartrate</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Diltiazem</td>
<td>1 (5)</td>
</tr>
<tr>
<td>None</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>11 (55)</td>
</tr>
<tr>
<td>Never smoker</td>
<td>9 (45)</td>
</tr>
<tr>
<td>CPAP(^b) usage, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (50)</td>
</tr>
<tr>
<td>Yes, but not compliant</td>
<td>3 (15)</td>
</tr>
<tr>
<td>No</td>
<td>7 (35)</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

This study presents a pilot effort to assess the level of agreement in HR measurements between PPG technology using the Apple Watch (1st generation) and telemetry during episodes of AF. We demonstrate that with a Lin’s concordance correlation coefficient of 0.88, the Apple Watch provided acceptable agreement with HR measurements by telemetry even during these episodes. The mean bias between the Apple Watch and telemetry measurements was 0.26 bpm, with 95% of Apple Watch HR measurements falling within 19 bpm of the telemetry measurements.

While the Lin’s concordance correlation coefficient is deemed accepted by the literature [18], we note that this interval was relatively wide, indicating there were relatively large differences in measurement. Furthermore, there appears to be an increase in variability of the differences as the magnitude of HR measurements increases, which casts doubt on the appropriateness of the constant mean bias assumption. While it is still subject to clinical judgment of how far apart HR measurements could be before 2 methods could be considered interchangeable, as Bland and Altman [19] note, the limits of agreement will be widened to some extent by the violation of the constant mean bias assumption, which thus would not lead to the acceptance of poor methods of measurement. As such, we adjusted for the average HR measurement in the mixed effects regression model to produce limits of agreement that better reflect the data [17,19].

Limitations

Our study is not without limitations. Despite screening 201 patients over a span of 6 months, only 35 patients were eligible, due to the criteria of having objective documentation of OSA. Furthermore, as this was a pilot study and to maximize yield of HR measurements while in AF, we aimed to enroll only 20 patients, yielding 134 HR measurements for each device (268 between the Apple Watch and telemetry) for analysis. Moreover, our small sample population was skewed toward white/Caucasian males. Because enrollment occurred in the preprocedure setting among patients who have established care with an electrophysiologist, the majority of participants demonstrated good rate control, and only 15% (n=3) were in RVR. This makes it difficult to assess the accuracy of PPG technology in measuring elevated HR and detecting periods of RVR, although our data support prior work suggesting that smart watches underestimate HR in these higher ranges [20,21]. Additionally, our data were collected under the direct supervision of a team member (RS), while the participants were sedentary, ensuring adequate skin contact between the smart watch and skin to obtain HR measurements. Generalizability of our results, therefore, may be limited and further studies with
a more diverse patient population and range of HR are needed, in sedentary and mobile settings. Furthermore, as our patient population was individuals with a known history of AF, our study did not demonstrate the ability to detect AF episodes, but rather the level of agreement on reported HR measurements with that of telemetry as the gold standard. This study frames the implications of our findings as an assessment of rate control rather than the actual detection of AF episodes. Regardless, we believe these data remain clinically useful for clinicians and patients aiming to evaluate adherence to treatment and titrate therapies accordingly.

Comparison With Prior Work
Because of its clinically silent nature, AF is difficult to detect, and guideline-directed management involves anticoagulation, rate control, and rhythm control [22]. A user-friendly device that allows for passive, noninvasive, and real-time HR monitoring, even during AF episodes, would therefore have substantial clinical implications for evaluating treatment efficacy. Smart watches and other wearables may be well-positioned to provide non-obtrusive, real-time HR monitoring and AF detection over long periods, limited only by battery life, wear time, and sensor algorithms.

Although several studies have evaluated the validity of smart watch algorithms to detect AF in healthy adults without cardiovascular disease [3,6,23], while some have assessed HR accuracy in wrist-worn monitors among healthy participants or patients with cardiovascular disease [24], our work adds to the body of research by showcasing promise regarding the accuracy of HR measurement via mobile health (mHealth) technology specifically in patients who are in AF. Thus, for individuals at high risk for AF—including those diagnosed with OSA, obesity, valvular disease, or hypertension [25]—smart watches and other wearables may serve as an important clinical tool. Furthermore, for patients who are diagnosed with OSA, passive HR monitoring may be particularly beneficial for noninvasive detection of AF. As previously noted, patients with OSA are at greater risk of AF recurrence after cardioversion, catheter ablation, and other antiarrhythmic therapies [7,10,26].

Moreover, by providing a larger cohort of data collected over a period in an ambulatory environment rather than within the restrictions of a clinic or hospital setting, smart watches have the potential to empower patients in their conversations with their health care providers regarding the efficacy of their AF therapies, including antiarrhythmic and rate control medications. This has been demonstrated in our clinical practice, where we have had patients with OSA self-identify an AF episode with RVR by a fast HR on their Apple Watch [24]. Our study may help clinicians understand the clinical utility of these ambulatory data should AF patients share the HR measurements from their Apple Watch. For patients with comorbid diagnoses of AF and OSA, the ability to passively monitor their HR with a smart watch may also promote adherence to OSA treatments including CPAP therapy and lifestyle modification, as these therapies have been shown to reduce AF recurrence and maintain sinus rhythm [26,27].

These patient–clinician conversations, informed by patient-generated data, could in turn promote adherence to guideline-directed management [28]. Current guidelines for the management of AF already address therapies including anticoagulation and rhythm control, risk factor modification (including OSA management), and remote device detection of AF through implantable devices [29]. Notably, the 2019 American Heart Association/American College of Cardiology/Heart Rhythm Society’s focused update to these guidelines remarked that “smart” or Wi-Fi-enabled devices may play a future role in the care of AF and be included in future recommendations [29]. As wearables continue to incorporate new technologies and the field of direct-to-consumer health informatics continues to evolve and address cardiovascular disease prevention and management, it is imperative that clinicians, researchers, and industry experts establish long-term collaborations to ensure that the products are accurate, safe, and beneficial without compromising clinical workflow or overwhelming the health care system.

Conclusions
In this study, we demonstrated that during AF episodes, HR readings from a commercially available smart watch (first-generation Apple Watch) are in acceptable agreement with HR measurements by telemetry, using patients with OSA as a proxy for a high-risk population. Further studies with larger sample populations and a wider range of HR are needed to confirm these findings. As ownership of smart devices and wearables continues to grow, our work demonstrates that these devices hold promise as tools to monitor efficacy of rate control therapies for patients with AF.

Acknowledgments
We thank each of the patients who participated in this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest
FM and SM are founders of and hold equity in Corrie Health, which intends to further develop the digital platform. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. Outside of this work, they have received material support from Apple and iHealth and funding from the Maryland Innovation Initiative, Wallace H. Coulter Translational Research Partnership, Louis B. Thalheimer Fund, the Johns Hopkins Individualized Health Initiative, and American Heart Association. SM also reports additional research support from the Aetna Foundation, the American Heart Association, the David and June Trone Family Foundation, Google, the National Institutes of Health, Nokia, and the PJ
References


**Abbreviations**

- AF: atrial fibrillation
- CPAP: continuous positive airway pressure
- HR: heart rate
- ICD: implantable cardioverter-defibrillator
- ILR: implantable loop recorder
- OSA: obstructive sleep apnea

©Pauline Huynh, Rongzi Shan, Ngozi Osuji, Jie Ding, Nino Isakadze, Françoise A Marvel, Garima Sharma, Seth S Martin. Originally published in JMIR Cardio (http://cardio.jmir.org), 08.02.2021. 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 http://cardio.jmir.org, as well as this copyright and license information must be included.
Development and Validation of an Automated Algorithm to Detect Atrial Fibrillation Within Stored Intensive Care Unit Continuous Electrocardiographic Data: Observational Study

Allan J Walkey¹, MD, MSc; Syed K Bashar², BSc; Md Billal Hossain², BSc; Eric Ding³, BA; Daniella Albuquerque³, BA; Michael Winter⁴, MSc; Ki H Chon², PhD; David D McManus³, MD

¹Boston University School of Medicine, The Pulmonary Center, Boston, MA, United States
²Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States
³University of Massachusetts Medical School, Worcester, MA, United States
⁴Boston University School of Public Health, Boston, MA, United States

Corresponding Author:
Allan J Walkey, MD, MSc
Boston University School of Medicine
The Pulmonary Center
75 East Concord Street
R-301
Boston, MA, 02118
United States
Phone: 1 617 638 4844
Email: allanjwalkey@gmail.com

Abstract

Background: Atrial fibrillation (AF) is the most common arrhythmia during critical illness, representing a sepsis-defining cardiac dysfunction associated with adverse outcomes. Large burdens of premature beats and noisy signal during sepsis may pose unique challenges to automated AF detection.

Objective: The objective of this study is to develop and validate an automated algorithm to accurately identify AF within electronic health care data among critically ill patients with sepsis.

Methods: This is a retrospective cohort study of patients hospitalized with sepsis identified from Medical Information Mart for Intensive Care (MIMIC III) electronic health data with linked electrocardiographic (ECG) telemetry waveforms. Within 3 separate cohorts of 50 patients, we iteratively developed and validated an automated algorithm that identifies ECG signals, removes noise, and identifies irregular rhythm and premature beats in order to identify AF. We compared the automated algorithm to current methods of AF identification in large databases, including ICD-9 (International Classification of Diseases, 9th edition) codes and hourly nurse annotation of heart rhythm. Methods of AF identification were tested against gold-standard manual ECG review.

Results: AF detection algorithms that did not differentiate AF from premature atrial and ventricular beats performed modestly, with 76% (95% CI 61%-87%) accuracy. Performance improved (P=.02) with the addition of premature beat detection (validation set accuracy: 94% [95% CI 83%-99%]). Median time between automated and manual detection of AF onset was 30 minutes (25th-75th percentile 0-208 minutes). The accuracy of ICD-9 codes (68%; P=.002 vs automated algorithm) and nurse charting (80%; P=.02 vs algorithm) was lower than that of the automated algorithm.

Conclusions: An automated algorithm using telemetry ECG data can feasibly and accurately detect AF among critically ill patients with sepsis, and represents an improvement in AF detection within large databases.

(JMIR Cardio 2021;5(1):e18840) doi:10.2196/18840

KEYWORDS
atrial fibrillation; sepsis; intensive care unit; big data; data science
Introduction

Atrial fibrillation (AF) is the most common arrhythmia during critical illness [1]. Among the most common causes of critical illness is sepsis—the potentially life-threatening syndrome caused by a dysregulated response to infection [2]. New-onset AF during sepsis is of special concern, as it is associated with increased mortality [3,4] and stroke risk [5], and likely represents a sepsis-defining organ dysfunction [6]. Despite the associated high morbidity and mortality, few studies have investigated potential mechanisms or optimal treatments of new-onset AF during sepsis. Given that large-scale manual review of continuous electrocardiographic (ECG) recordings is not feasible, and administrative data do not allow identification of AF timing, there has been increasing interest in developing and refining automated algorithms for the detection of AF in electronic health record data that facilitate AF research [7]. However, automated AF detection among critically ill patients with sepsis faces additional challenges, including telemetry data that may be subject to high burdens of premature beats, other arrhythmias, noise [8], and signal loss. Reliable, real-time, automated approaches to accurately identify ECG noise and artifacts are critical to accurate identification of AF in an intensive care unit (ICU) setting and are underdeveloped. We sought to (1) develop, validate, and iteratively evaluate the performance of a novel algorithm that incorporates the critical elements necessary for AF identification during critical illness including noise elimination, premature atrial and ventricular beat detection [9], and AF detection, using a large-scale, electronic health database with standard telemetry ECG data, and (2) compare performance characteristics of automated AF identification with other methods of AF ascertainment within electronic health record data.

Methods

Cohort

We identified adult patients with sepsis defined by ICD-9 (International Classification of Diseases, 9th edition) codes for infection and acute organ dysfunction as described previously [10] using Medical Information Mart for Intensive Care (MIMIC III) open source medical record data [11]. MIMIC III is a single-center database from a large tertiary care hospital, with linked ECG telemetry waveform and electronic medical record information from patients hospitalized between 2001 and 2012. Patients without a linked waveform file, with a paced rhythm, with absent or corrupted ECG recordings, with fewer than 6 hours of ECG telemetry data, or with more than 55 hours of ECG telemetry data were excluded from the analysis.

Waveform Selection and Gold-Standard Rhythm Status Determination

We performed iterative training and testing of automated AF detection algorithms. We selected 25 candidate case patients with AF during sepsis and 25 candidate control patients without AF during sepsis as identified by ICD-9 codes (427.31). The 50 candidate waveforms were then reviewed manually by trained study staff (DA and ED) with the final adjudication of rhythm status (sinus rhythm vs AF) by a board-certified clinical cardiac electrophysiologist (DM) as the gold standard [12]. The 50 candidate waveforms were sent to the algorithm development team for adjudication of rhythm status via the automated algorithm. Investigators involved with algorithm development and testing (MH, SB, and KC) were blinded to each patient’s gold-standard rhythm determination (sinus rhythm or AF).

Automated AF Detection Algorithm

Continuous telemetry ECG recordings between 6 and 55 hours in length and with at least one readable ECG recording were divided into 2-minute segments, which were first analyzed for interpretable signal using automated signal and noise detection [13]. The 2-minute ECG segments without a predominance of noise were then analyzed with a novel R-wave detection method that detects QRS complexes using variable-frequency complex demodulation–based ECG reconstruction [14]. Next, the variability of R–R intervals was evaluated using sample entropy, a measure of randomness that is expected to be higher for patients with AF than those with normal sinus rhythm [15]. Based on the sample entropy calculated from the R–R intervals, an automated “initial screening” for AF was performed, where the “possible AF” status may include premature atrial and ventricular contraction segments as false-positive detections of AF. In order to differentiate increased R–R randomness from AF in contrast to R–R variability caused by premature atrial and ventricular beats, a novel premature beat detection step was added to the algorithm which only takes the “possible AF” segments determined by the sample entropy in the previous step [16]. Two approaches were used to differentiate premature atrial and ventricular beats from AF. First, Poincaré plots derived from the differences of heart rates were used to differentiate AF from premature atrial and ventricular beats as repeated triangular-shaped patterns were found for premature atrial and ventricular contractions in the Poincaré plot [9]. In addition to the Poincaré plots, P-waves were identified using a recently developed empirical mode decomposition–based algorithm [17]. Because AF is characterized by an absence of P-waves, but premature atrial and ventricular beats occur in the midst of sinus rhythms with P-waves that precede QRS complexes, high ratios of P-wave to R-wave were used to aid differentiation of premature beats from AF (low P-to-R ratio) [16]. Further, in order to increase the specificity of the AF detection algorithm, we a priori determined that the automated AF detection algorithm would identify a patient as having an AF episode only if 3 consecutive 2-minute ECG segments (6 minutes) were identified as containing continuous AF. The algorithm identified AF in one of the ECG leads, though an exploratory post-hoc analysis made all ECG leads available to the automated algorithm. A summary of the AF detection algorithm is shown in Figure 1.
AF Algorithm Development and Validation

AF detection algorithms were derived and validated in a stepwise manner (Figure 2). The AF detection algorithms using only automated noise detection and R–R sample entropy were first trained using selected waveforms without premature beats (training set 1, Round 1) and then validated (test set 1, Round 2) using randomly selected waveforms with and without AF. In order to determine the added value of premature beat detection, we added automated premature atrial and ventricular beat detection using Poincaré plots, and then added P-to-R-wave ratios to the algorithms tested in Rounds 1 and 2 and retested the algorithm in test set 1. In the final validation experiments (test set 2), we deployed the complete ensemble algorithm, which included noise detection, R–R sample entropy, and premature atrial and ventricular beat detection with Poincaré and P-wave detection, using 50 randomly chosen AF and non-AF waveforms. In total, 3 cohorts with 150 patients were evaluated using manual AF detection with results blinded to the deployment of the automated algorithm.

**Statistical Analyses**

We evaluated agreement between the gold-standard review of telemetry ECG data by an expert ECG reader (DM) and other methods of AF detection including the automated AF detection algorithm, nurse charting of AF status, and ICD-9 codes using $2 \times 2$ contingency tables. Additionally, we performed a post-hoc
exploratory analysis to evaluate the performance of previously
described automated methods of AF detection in our test set—a
statistical method [17] that used the root mean square of
successive differences, Shannon entropy, and turning point ratio
calculated from R–R intervals to automatically detect AF; and
a method [18] that used the coefficient of sample entropy
obtained from R–R intervals to determine the AF status.

Sensitivity (true-positive rate), specificity (true-negative rate),
positive (proportion of positive signals that are true positives)
and negative predictive values (proportion of negative signals
that are true negatives) were calculated for each AF algorithm
with 95% confidence intervals using MedCalc (MedCalc
Software). We calculated the average time between estimates
of AF onset for the gold standard as compared with other

Comparisons of accuracy were conducted with \( \alpha = .05 \). All study
procedures were deemed not human subjects research by the
Boston University Medical Campus and University of
Massachusetts Medical School Institutional Review Boards.

**Results**

Among 58,976 ICU admissions within MIMIC III, we identified
14,831 admissions for adults with sepsis, among whom 2975
patients had ECG waveforms linked to clinical data. Three
groups of 50 ECG waveforms from patients hospitalized with
sepsis were randomly selected and evaluated iteratively through
the automated AF detection algorithm. Characteristics of patients
with waveforms selected for algorithm validation (test set 2)
are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Patient characteristics validation cohort (N=50).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td>Age, mean (SD)</td>
</tr>
<tr>
<td>Sex, female, n (%)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>Black, n (%)</td>
</tr>
<tr>
<td>White, n (%)</td>
</tr>
<tr>
<td>Asian, n (%)</td>
</tr>
<tr>
<td>Other, n (%)</td>
</tr>
<tr>
<td>Unknown, n (%)</td>
</tr>
<tr>
<td>Comorbidity score (Elixhauser–van Walraven), mean (SD)</td>
</tr>
<tr>
<td>Heart failure, n (%)</td>
</tr>
<tr>
<td>Coronary artery disease, n (%)</td>
</tr>
<tr>
<td>Valvular disease, n (%)</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
</tr>
<tr>
<td>Diabetes without complication, n (%)</td>
</tr>
<tr>
<td>Diabetes with complication, n (%)</td>
</tr>
<tr>
<td>Chronic pulmonary disease, n (%)</td>
</tr>
<tr>
<td>Renal failure, n (%)</td>
</tr>
<tr>
<td>Obesity, n (%)</td>
</tr>
<tr>
<td>Source of infection</td>
</tr>
<tr>
<td>Pulmonary infection, n (%)</td>
</tr>
<tr>
<td>Genitourinary infection, n (%)</td>
</tr>
<tr>
<td>Gastrointestinal infection, n (%)</td>
</tr>
<tr>
<td>Skin/Soft tissue infection, n (%)</td>
</tr>
<tr>
<td>Cardiovascular infection, n (%)</td>
</tr>
<tr>
<td>Postprocedural infection, n (%)</td>
</tr>
<tr>
<td>Unspecified septicemia, bacteremia, n (%)</td>
</tr>
<tr>
<td>Other infection, n (%)</td>
</tr>
<tr>
<td>SOFA(^a) score, ICU(^b) day 1, mean (SD)</td>
</tr>
</tbody>
</table>

\(^a\)SOFA: Sequential organ failure assessment.
\(^b\)ICU: intensive care unit.
Results of the rounds of training and testing (validation) of the AF detection algorithm are shown in Tables 2-7. During the initial training round on waveforms selected for lack of noise, the accuracy of the R–R sample entropy algorithm for AF detection was 100% (Table 2). However, the algorithm performed more modestly when deployed in a separate set of unselected ECG waveforms (accuracy 76%; 95% CI 61%-87%; Table 3). Analysis of potential reasons for discrepancy between the automated algorithm and manual abstraction showed that the lower specificity and positive predictive value from the algorithm were due to false-positive AF detection in the setting of critically ill patients with a high burden of premature atrial beats. We next trained an algorithm to discriminate between premature beats and AF, first using Poincaré analysis (Table 4, 94% accuracy), and then with a combination Poincaré and P-wave detection (Table 5, 98% accuracy), both of which demonstrated improvement in accuracy (P=.02) as compared with models without the ability to detect premature beats. The high accuracy (94%, 95% CI 83%-99%) and low false-positive rate (96% specificity) were confirmed in a separate validation cohort (test cohort 2; Table 6). Exploratory analysis showed further improvement in algorithm performance with analysis of all available ECG leads for each patient (Table 7, 98% accuracy). The time of AF onset detected by the automated algorithm differed from the manual detection onset time by a median 30 minutes (25th-75th percentile 0-208 minutes), potentially related to the algorithm’s requirement for 3 consecutive 2-minute segments to be present to call AF.

Table 2. Results of training set[^a] for initial automated atrial fibrillation identification using signal and noise detection and R–R interval indices, from selected waveforms without premature beats.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>No atrial fibrillation</th>
<th>Atrial fibrillation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td></td>
<td>24</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td></td>
<td>0</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
<td>25</td>
<td>49[^b]</td>
</tr>
</tbody>
</table>

[^a] The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 100% (95% CI 86%-100%), 100% (95% CI 86%-100%), 100%, 100%, and 100% (95% CI 93%-100%), respectively.

[^b] One patient had noise in the waveform tracings that were unable to be run through the algorithm due to lack of 3 consecutive 2-minute segments of majority noise-free time.

Table 3. Results of test set[^a] of automated atrial fibrillation detection using signal and noise detection and R–R interval indices only, from unselected waveforms.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>No atrial fibrillation</th>
<th>Atrial fibrillation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td></td>
<td>10</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td></td>
<td>1</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11</td>
<td>38</td>
<td>49[^b]</td>
</tr>
</tbody>
</table>

[^a] The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 91% (95% CI 59%-100%), 71% (95% CI 54%-85%), 48% (95% CI 35%-61%), 96% (95% CI 80%-99%), and 76% (95% CI 61%-87%), respectively.

[^b] One patient had extensive noise in the waveform tracings that were unable to be run through the algorithm.

Table 4. Results of test set[^a] for automated atrial fibrillation using signal and noise detection and R–R interval indices, with added Poincaré plots to detect premature beats from unselected waveforms.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>No atrial fibrillation</th>
<th>Atrial fibrillation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td></td>
<td>11</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td></td>
<td>0</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11</td>
<td>38</td>
<td>49</td>
</tr>
</tbody>
</table>

[^a] The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 100% (95% CI 72%-100%), 90% (95% CI 75%-97%), 73% (95% CI 52%-87%), 100%, and 92% (95% CI 80%-98%), respectively.
Table 5. Results of test set 1\(^a\) for automated atrial fibrillation using signal and noise detection and R–R interval indices, with added Poincaré, and P-wave indices to detect premature beats from unselected waveforms.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>49</td>
</tr>
</tbody>
</table>

\(^a\)The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 100% (95% CI 72%–100%), 95% (95% CI 82%–99%), 85% (95% CI 59%–96%), 100%, and 96% (95% CI 86%–99%), respectively.

Table 6. Results of the test set 2\(^a\) for automated atrial fibrillation signal and noise detection and R–R interval indices, with added P-wave and premature beat detection, from unselected waveforms.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

\(^a\)The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 92% (95% CI 74%–99%), 96% (95% CI 80%–100%), 96% (95% CI 77%–99%), 92% (95% CI 76%–98%), and 94% (95% CI 83%–99%), respectively.

Table 7. Exploratory analysis on test set 2\(^a\), evaluating effect of using all available leads (rather than only 1 lead shown in for automated atrial fibrillation detection.

<table>
<thead>
<tr>
<th>Automated algorithm atrial fibrillation status</th>
<th>Manual atrial fibrillation status</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

\(^a\)The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 100% (95% CI 86%–100%), 96% (95% CI 80%–99.9%), 96% (95% CI 79%–99%), 100% (95% CI 76%–98%), and 98% (95% CI 89%–99.9%), respectively.

We also compared administrative codes (ICD-9; Table 8) and nurse cardiac rhythm annotation (Table 9) with manual AF identification. Compared with the gold-standard manual review of the ECG telemetry data, ICD-9 codes associated with AF showed 68% agreement (P=.002 vs automated algorithm) and nurse annotation of AF showed 80% agreement (P=.02 vs automated algorithm). Although timing of AF onset could not be estimated from ICD-9 codes, nurse charting of AF onset occurred a median of 56 minutes after gold-standard AF onset time (25th to 75th percentile of –13 to 705 minutes).

Table 8. Comparison of manual electrocardiographic detection of atrial fibrillation with ICD-9 codes for atrial fibrillation.\(^a\)

<table>
<thead>
<tr>
<th>ICD-9(^b) codes</th>
<th>Manual atrial fibrillation status</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atrial fibrillation</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

\(^a\)The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 64% (95% CI 43%–82%), 72% (95% CI 50%–88%), 70% (95% CI 53%–82%), 67% (95% CI 53%–78%), and 68% (95% CI 53%–80%), respectively.

Finally, we performed an exploratory analysis to evaluate the performance of 2 previously described AF detection algorithms [18,19] (Multimedia Appendices 1 and 2). Both previously described algorithms had numerically lower accuracy than our novel AF-detection ensemble (Dash et al [18]: 90% accuracy, 95% CI 78%–97%; Lake and Moorman [19]: 76% accuracy, 95% CI 62%–87%).

**Discussion**

**Principal Findings**

We developed, validated, and evaluated a novel, automated, accurate algorithm to detect AF from stored electronic health record ECG waveform data from telemetry recordings. We used a stepwise approach to algorithm development and demonstrated that the automated AF detection algorithm worked by first eliminating waveforms with noisy segments that impairs reliable rhythm assessment, next by discriminating premature atrial and ventricular beats that mimic the rhythm irregularity from AF, and finally by using R-wave variability algorithms to detect AF from 2-minute-long ECG segments. The automated algorithm demonstrated predictive values greater than 90% and detecting AF within a median 30 minutes of manual ascertainment. The automated algorithm showed favorable performance characteristics when compared with currently available standard methods of large-scale AF ascertainment, including diagnostic codes, nurse annotation of rhythm status recorded in the electronic medical record, and previously described automated AF detection approaches [18,19].

**Limitations**

Our findings should be considered in light of study limitations. Data arose from a single center and diagnostic claims coding and nurse documentation of heart rhythm status may differ at other centers. Further testing of the performance of the automated AF detection algorithm in other settings and in comparison to other automated methods of AF detection, such as machine learning techniques, is certainly warranted. Strengths of this study include the manual validation of all key ECG segments by trained study personnel with oversight of an expert ECG reader, use of an algorithm that automates signal and noise detection, and the stepwise analysis quantifying improvement in algorithm performance when adding different features, which demonstrate the necessity of adding premature beat detection to an algorithm designed to detect AF in the setting of critical illness.

**Comparison With Prior Work**

Few prior studies have evaluated automated algorithms for AF detection among critically ill patients. Moss et al [7] tested an algorithm using an ensemble of R–R interval time-series approaches previously developed from outpatient Holter rhythm monitoring [19] among 500 30-minute telemetry segments of ICU patients in a single center, and found sensitivity and positive predictive value of 89% and 99%, respectively. The method of AF detection by Moss et al [7] differed from our algorithm in multiple ways: we used automated noise detection to select evaluable ECG segments, required shorter ECG segments for analysis (2 minutes vs 10 minutes), and combined R–R time-series approaches (ie, sample entropy and Poincaré plot features) with P-wave characteristics in order to discriminate premature beats from AF. Accuracy of AF onset times were not reported in the Moss et al [7] ICU sample. Although we do not directly compare the algorithm described by Moss et al using MIMIC ECG data, use of an earlier iteration of the ensemble used by Lake and Moorman [19] showed less favorable accuracy within our cohort when compared with our novel algorithm. Results from our stepwise, iterative analysis of automated algorithm performance demonstrated the importance of incorporating strategies that could identify P-waves and differentiate premature atrial and ventricular beats from AF among critically ill patients with sepsis. Given differences in patient characteristics and validation strategies between Moss et al [7] and our study, further studies comparing different automated approaches to AF detection within an independent validation cohort are warranted.

In addition to determining accuracy of a novel, automated ECG detection algorithm for AF detection, we also evaluated existing methods of AF recognition within claims data ICD-9 codes and electronic medical record–based nurse annotation of heart rhythm. Compared with manual ECG review, ICD-9 codes were unable to identify AF timing and showed only modest performance (68% accuracy, 70% positive predictive value, and 67% negative predictive value) for correctly identifying cases of AF during the ICU stay. Nurse charting of heart rhythm status performed similar to ICD-9 codes for rhythm status determination, and although nurse charting allowed for timing of AF episodes [12], AF onset times from nurse-charted AF episodes differed from the gold-standard rhythm onset by approximately 1 hour. Thus, in our sample of patients with sepsis, automated AF detection was superior to current standard large-scale approaches to AF detection using electronic health record data. Prior studies validating ICD-9 codes for AF ascertainment had lower predictive values than our findings, specifically from 502525Total

<table>
<thead>
<tr>
<th>Nurse charting</th>
<th>Manual atrial fibrillation status</th>
<th>No atrial fibrillation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>24</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>No atrial fibrillation</td>
<td>1</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

(a) The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 96% (95% CI 80%–100%), 64% (95% CI 43%–82%), 73% (95% CI 61%–82%), 94% (95% CI 70%–99%), and 80% (95% CI 66%–90%), respectively.
detection showed better performance than our sample [5], potentially because ECG data were available only from ICU in this study, rather than the entire hospitalization.

Multiple potential uses exist for an algorithm that can accurately read and identify AF from ECG waveform data from critically ill patients with sepsis. Our automated AF detection algorithm is a novel tool that facilitates the analysis of underutilized continuous waveform data currently housed in electronic data repositories, and allows AF to be studied using “big data” analytic approaches. Large-scale AF identification can be used in future studies to evaluate risk factors and triggers of AF, and to study long-term ramifications of subclinical AF occurring during acute illness such as sepsis. Because of the automated detection of ECG signal, noise, premature beats, and AF, the algorithms can also be adapted and scaled for rapid, real-time identification of AF among patients undergoing continuous ECG monitoring, including critically ill patients with complex ECG waveforms. The AF algorithm based on sample entropy is computationally more efficient than machine learning algorithms that require significant training data, and reports similar accuracy to machine learning methods not subjected to the additional challenge of high premature beat burdens met by the present algorithm among critically ill patients [20-22]. Furthermore, algorithm development was hypothesis driven, enabling us to understand the relative contributions of premature beats and ECG noise to overall AF detection performance. Despite the fact that the prevalence of AF in unselected ambulatory populations may be lower than in our sample of inpatients with sepsis, our AF detection approach with noise cancellation and premature beat discrimination may also be useful in ambulatory ECG data from Holter monitors [23] and ECG data from wearable devices, as these devices are also frequently affected by motion and noise artifact.

Conclusions

We derived and validated an automated algorithm that detects an ECG signal, eliminates segments corrupted by noise artifact, and can discriminate AF from other causes of irregular R–R intervals such as premature atrial and ventricular beats. The automated algorithm performed with higher accuracy than currently available methods for large-scale AF detection, including ICD codes and nurse charting of heart rhythm status from data in the electronic health record. Further studies can use the algorithm to identify AF in large-scale electronic health record data to facilitate studies of risk factors and triggers of AF, as well as long-term complications of subclinical AF during acute illness.

Acknowledgments

AW and the work under consideration were funded through NIH NHLBI R01HL136660; ED is funded through NIH grant 5T32HL120823. DM receives research support from grants U54HL143541, R01HL126911, R01HL137734, R01HL137794, and R01HL135219 from the National Heart, Lung, and Blood Institute and National Center for Complementary and Integrative Health (National Institutes of Health), and Grant NSF-12-512 from the National Science Foundation. DM has also received research support from Bristol Myers Squibb, Care Evolution, Samsung, Apple, Pfizer, Biotronik, Boehringer Ingelheim, Philips Research Institute, and has consulted for Bristol Myers Squibb, Pfizer, Philips, Samsung Electronics, and FLEXCon.

Conflicts of Interest

DM received grants, research support and/or consulting fees from Bristol Myers Squibb, Pfizer, Flexcon, Samsung, Philips, Apple, Boeringher Ingelheim, Fitbit, NIH, NSF, Heart Rhythm Society, Rose Consulting, Boston Biomedical Associates.

Multimedia Appendix 1

Results of an alternate automated atrial fibrillation detection method (Dash et al [18]; Test set 2).

[DOCX File, 13 KB - cardio_v5i1e18840_app1.docx ]

Multimedia Appendix 2

Comparison of automated atrial fibrillation detection with COSEn-based method (Lake and Moorman [19]; Test set 2).

[DOCX File, 13 KB - cardio_v5i1e18840_app2.docx ]

References


Abbreviations

AF: atrial fibrillation
Internet-Delivered Exposure-Based Therapy for Symptom Preoccupation in Atrial Fibrillation: Uncontrolled Pilot Trial

Josefin Särnholm, MSc; Helga Skúladóttir, MD; Christian Rück, PhD, MD; Sofia Klavebäck, MD; Eva Ölafsdóttir, RNC; Susanne S Pedersen, PhD; Frieder Braunschweig, PhD, MD; Brjánn Ljóttson, PhD

Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet, Stockholm, Sweden
Department of Medicine, Cardiology Unit, Karolinska Institutet, Stockholm, Sweden
Department of Clinical Neuroscience, Karolinska Institutet, Centre for Psychiatry Research, Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden
Department of Medicine, Cardiology Unit, Karolinska Institutet, Department of Cardiology, Karolinska University Hospital, Stockholm, Sweden
Department of Psychology, University of Southern Denmark, Department of Cardiology, Odense University Hospital, Odense, Denmark

Corresponding Author:
Josefin Särnholm, MSc
Department of Clinical Neuroscience
Division of Psychology
Karolinska Institutet
Nobelsv. 9
Stockholm, 171 65
Sweden
Phone: 46 736373295
Email: josefin.sarnholm@ki.se

Abstract

Background: Atrial fibrillation (AF) is the most common cardiac arrhythmia in the adult population. AF is associated with a poor quality of life (QoL) and, in many patients, current medical treatments are inadequate in alleviating AF symptoms (eg, palpitations). Patients often present with symptom preoccupation in terms of symptom fear, avoidance, and control behaviors. Internet-delivered cognitive behavior therapy is effective for treating other somatic disorders but has never been evaluated in patients with AF.

Objective: The aim of this study is to evaluate the efficacy and feasibility of AF-specific internet-delivered cognitive behavior therapy.

Methods: We conducted an uncontrolled pilot study in which 19 patients with symptomatic paroxysmal AF underwent internet-delivered cognitive behavior therapy. Participants completed self-assessments at pretreatment, posttreatment, and at a 6-month follow-up along with handheld electrocardiogram measurements with symptom registration. The treatment lasted 10 weeks and included exposure to physical sensations, reduction in avoidance behavior, and behavioral activation.

Results: We observed large within-group improvements in the primary outcome, AF-specific QoL (Cohen d=0.80; P<.001), and in symptom preoccupation (Cohen d=1.24; P<.001) at posttreatment; the results were maintained at the 6-month follow-up. Treatment satisfaction and adherence rates were also high. We observed an increased AF burden, measured by electrocardiogram, at the 6-month follow-up, but a significant decrease was observed in the overestimation of AF symptoms at posttreatment and 6-month follow-up. Exploratory mediation analysis showed that a reduction in symptom preoccupation mediated the effects of internet-delivered cognitive behavior therapy on AF-specific QoL.

Conclusions: This study presents preliminary evidence for the potential efficacy and feasibility of a novel approach in treating patients with symptomatic AF with internet-delivered cognitive behavior therapy.

Trial Registration: ClinicalTrials.gov NCT02694276; https://clinicaltrials.gov/ct2/show/NCT02694276

(JMIR Cardio 2021;5(1):e24524) doi:10.2196/24524

KEYWORDS
atrial fibrillation; arrhythmia; cognitive behavior therapy; quality of life; anxiety
Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, with a prevalence of approximately 3% in the general adult population [1]. AF is caused by disturbances in the heart’s electrical signaling, resulting in irregular and often rapid heartbeats, and is associated with a wide range of symptoms such as palpitations, dizziness, fatigue, chest pain, and dyspnea [2]. AF episodes often have a sudden onset and may last briefly or for days; moreover, AF is classified as paroxysmal if the conversion back to the normal sinus rhythm occurs spontaneously within a week [3]. Current treatment strategies (pharmacological and invasive therapies) are associated with potentially serious side effects and do not sufficiently reduce the symptom burden in many patients with AF [4,5]. AF is associated with a poor quality of life (QoL) [6], an increased risk for developing anxiety disorders and depression [7], and high rates of hospitalization and health care utilization, further leading to high societal costs [8]. The QoL impairment in AF has been shown to be unrelated to the objectively measured arrhythmia burden (frequency and duration of AF episodes) [9] and objective measures of disease severity (eg, cardiac dysfunction) [10].

AF burden can be defined as either objective AF episodes, measured and confirmed by an electrocardiogram (ECG), or subjectively perceived AF symptoms measured by self-report [11]. When comparing ECG-recorded episodes of AF and patients’ perceptions of symptoms, patients have been shown to both over- and underestimate the actual burden of AF [11]. Psychological distress has been shown to be predictive of overestimating AF symptoms, that is, indicating AF symptoms while in a normal cardiac rhythm [11-13], a discordance that may be explained by the presence of symptom preoccupation in patients with AF.

We have defined symptom preoccupation in AF as the fear of experiencing and triggering AF episodes, hypervigilation toward cardiac symptoms, persistent worry about complications (eg, stroke), and avoidance of physical and social activities [14]. Symptom preoccupation has shown to be a strong predictor of higher self-reported symptom severity, poor mental and physical QoL, and psychological distress in AF [15] and other somatic and functional disorders [16]. There are several possible pathways through which symptom preoccupation may impact clinical outcomes in AF. Fear and hypervigilation of cardiac-related symptoms may increase the likelihood that patients with AF misinterpret the normal cardiac activity or cardiac activity related to stress as arrhythmia. The perception of AF symptoms may trigger anxiety, leading to autonomic arousal and increases in heart rate, which can trigger extra beats and potential AF episodes [17]. When anticipating or experiencing AF symptoms, behavioral responses—such as avoidance behavior, symptom control behaviors, and excessive worry—may increase the awareness of symptoms [18] and maintain the fearful responses to AF symptoms [19], leading to more AF-related disability. Cardiac anxiety, driven by avoidance behavior, has been shown to affect clinical outcomes by increasing the risk of adverse cardiac prognosis in other heart diseases [20]. In summary, current evidence indicates that psychological and behavioral factors influence clinical outcomes in AF and that AF disability extends beyond the actual arrhythmia.

Cognitive behavior therapy (CBT), with an emphasis on exposure, aims to break the cycle of avoidance behavior, symptom fear, and disability through systematic contact with a stimulus that evokes conditioned aversive responses while abstaining from avoidance and safety behaviors [21]. Exposure therapy also encourages patients to willingly expose themselves to aversive stimuli by simultaneously engaging in a behavior that is inconsistent with the emotional response elicited by that stimuli [22]. Our research group has developed and evaluated an AF-specific CBT protocol. In a previous study, we conducted an investigation of exposure-based AF-specific CBT in a face-to-face format targeting symptom preoccupation. It showed promising effects on AF-specific QoL and self-reported AF symptoms [14].

One development in the effort to bridge the gap between the demand and availability of CBT [23] has been the delivery of treatment via the internet. Internet-CBT has been evaluated in a range of trials for somatic and psychiatric disorders, with large treatment effects [24], but has never been evaluated for AF. Studies on internet-CBT in patients with cardiac disorders are scarce [25], and there has been a call to increase access to digital health solutions in the cardiac health care community [26].

The purpose of this study is to evaluate the efficacy and feasibility of internet-delivered AF-specific CBT, primarily based on exposure exercises in preparation for a forthcoming randomized controlled trial (RCT). Additional objectives of the study are to investigate changes over time in the objectively measured AF burden and in the participants’ potential overestimation of their AF symptoms as well as to explore whether symptom preoccupation was a potential mediator of treatment effect on AF-specific QoL.

Methods

Design Overview

This is an uncontrolled pilot study with a pretest-posttest design and a 6-month follow-up. We aimed to recruit 30 participants to achieve a power of 80% for detecting a moderate improvement, as indicated by an effect size of Cohen $d=0.65$, in the primary outcome measure of disease-specific QoL between pre- and posttreatment assessments. However, only 19 participants were recruited because of a slow recruitment rate. An interim analysis after 14 participants had been recruited did not show any significant improvement in the primary outcome, and the discontinuation of recruitment was justified.

Participants

The participants were referred by cardiologists in Stockholm, Sweden. To be eligible for this study, participants had to have (1) paroxysmal AF as defined above and (2) receive optimal medical care according to the current clinical guidelines [3]. The following inclusion criteria also had to be fulfilled: (1) ≥1 AF episode per month and symptoms that troubled the patient or caused limitations in daily activities (ie, European Heart...
Rhythm Association [EHRA] class ≥IIb; [27]); (2) age 18-75 years; and (3) ability to read and write in Swedish. The exclusion criteria were as follows: (1) heart failure with severe systolic dysfunction (ejection fraction ≤ 35%); (2) significant valvular disease; (3) planned ablation for AF or ablation during the 3 months before assessment; (4) other severe medical illness; (5) any medical restriction on physical exercise; (6) severe psychiatric disorder, severe depression, or risk of suicide; and (7) alcohol dependency. Participants were asked not to participate in any concurrent psychological treatment during the course of this treatment and were encouraged, along with their referring cardiologists, to avoid changes in medical therapy unless clinically necessary. The study was approved by the regional ethical review board in Stockholm, Sweden (reg.no: 2015/1843-31/2) and was registered on ClinicalTrials.gov (NCT02694276). This report of the study adheres to the TREND Statement checklist for nonrandomized interventions [28].

Instruments

The primary outcome was the Atrial Fibrillation Effect on Quality-of-Life (AFEQT), which is an AF-specific measure of QoL that evaluates the following parameters: AF symptoms, impairment in social and physical activities, medical treatment concerns, and AF-specific treatment satisfaction [29]. The scale has 20 items, with the total score ranging from 0 (severe symptoms and disability) to 100 (no symptoms and disability). The total AFEQT score corresponded to the following categories of AF severity: mild (71.3, SD 20.6), moderate (57.9, SD 19.0), and severe (42.0, SD 21.2) [29]. The subscale AF-specific treatment satisfaction (two items) measures satisfaction with the current medical treatment and is not included in the total score; thus, it was not analyzed in this study.

Secondary outcome measures included the Symptoms Checklist, Frequency, and Severity Scale (SCL), which is a disease-specific checklist used to measure AF-related symptoms on two subscales - symptom severity and symptom frequency [30]. The University of Toronto Atrial Fibrillation Severity Scale was used to assess AF-specific symptoms [31] and AF-specific health care utilization. From that scale, we combined items 10 (visits to emergency room), 11 (number of hospitalizations), and 12 (visits to cardiologists) to assess cardiac-specific health care consumption. The Cardiac Anxiety Questionnaire (CAQ) measures cardiac anxiety using 3 subscales: (1) fear (eg, When my heart is beating fast, I get frightened), (2) avoidance (eg, I avoid exercise or other physical work), and (3) attention (eg, I pay attention to my heartbeat) [32] with a greater score indicating an elevated cardiac anxiety.

The 4-item Perceived Stress Scale (PSS-4) was used to assess stress sensitivity, [33] the 7-item Generalized Anxiety Disorder (GAD-7) scale was used to measure general worry and anxiety [34], and the Patient Health Questionnaire (PHQ-9) was used to measure depressive symptoms [35]. Participants also completed the World Health Organization Disability Assessment Schedule (WHODAS), a well-validated measure of general health and disability [36].

The Client Satisfaction Questionnaire (CSQ-8) was used to measure satisfaction with the CBT treatment [37]. At posttreatment and 6-month follow-up, we assessed whether participants had experienced any adverse events caused by their participation in the treatment. Participants were asked to report and rate the short- and long-term discomfort of the adverse event on a scale of 0 (did not affect me at all) to 3 (affected me very negatively) [38]. Other measures were also included, but these are not presented in this paper.

ECG Measurements

To assess the objective AF burden, participants were asked to perform a 30-second intermittent handheld ECG (Zenicor EKG thumb, Stockholm, Sweden) registration at home, four times daily, and while experiencing cardiac symptoms over 2 weeks at pre- and posttreatment and at 6-month follow-up. The participants were instructed to push a symptom indicator button on the device when they recognized that they had AF symptoms, both during the regular daily measurements and while experiencing cardiac symptoms. A more detailed technical description of the device is available elsewhere [39].

Intervention

The AF-specific exposure-based internet-CBT treatment was delivered completely via the internet through a tailored and secure web-based platform. The treatment was therapist-guided, lasted for 10 weeks, and included 5 interactive treatment modules with weekly homework assignments to be completed during the first 5 weeks. The modules were downloadable as PDF files and comprised between 13 and 16 pages (A4) of text, and a total of 68 pages. After the fifth module, participants continued to work with the treatment content for the remaining 5 weeks and sent weekly reports about their exposure exercises to their treating psychologist. Participants were encouraged to work with the treatment for approximately 30-60 minutes per day.

The treatment protocol was based on an AF-specific CBT protocol that was previously evaluated in a face-to-face pilot study [14] and an exposure-based CBT manual for irritable bowel syndrome [38]. The treatment primarily targets two maintenance processes of AF disability: hypervigilance and fear of AF symptoms and the avoidance of physical and social activities. The treatment comprised the following interventions: (1) education on AF (pathophysiology and medical treatment) and symptom preoccupation; (2) a self-observation exercise (ie, awareness of current cardiac symptoms, thoughts, feelings, and behavioral impulses to reduce the negative valence of symptoms); (3) exposure to physical sensations similar to AF symptoms by engaging in a variety of physical exercises, such as increasing the heart rate by running up and down the stairs or inducing dizziness or dyspnea by overbreathing to reduce symptom-related fear and hypervigilance; (4) in vivo exposure to avoided situations or activities where symptoms are unwanted (such as participating in a social activity while experiencing cardiac symptoms); (5) reduction or removal of behaviors, such as postponing pulse checking, that serve to control symptoms. Interventions (2)-(6) were used in conjunction to maximize their effectiveness, for example, participants were encouraged to use the self-observation exercise during exposure and to enhance exposure by inducing physical sensations while conducting exposure in vivo. AF episodes that occurred during treatment were also framed as an opportunity to practice the new skills.
acquired via the treatment. Examples of AF-related avoidance and control behaviors and how these could be targeted with exposure exercises were also presented in comprehensive clinical vignettes; (7) behavioral activation, where patients were encouraged to take steps toward goals within important life areas impaired by AF; and (8) in the final module, participants worked with relapse prevention to handle the potential progression of AF. The treatment is described in more detail in the report of our previous pilot study [14]. Textbox 1 provides an overview of the treatment components.

Textbox 1. Overview of treatment components.

<table>
<thead>
<tr>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pathophysiology of atrial fibrillation (AF).</td>
</tr>
<tr>
<td>• The role of anxiety and behavior on symptoms and quality of life.</td>
</tr>
<tr>
<td>• Brief training in self-observation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interoceptive exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Exposure to physical sensations similar to AF symptoms.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure in vivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Gradual exposure to avoided situations or activities that patients fear may trigger AF symptoms or where symptoms are unwanted.</td>
</tr>
<tr>
<td>• Combining in vivo exposure with interoceptive exposure.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Identifying life areas impaired by AF-related disability or symptom fear.</td>
</tr>
<tr>
<td>• Set behavioral goals and gradually take steps toward them.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relapse prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Prevention of relapse into control or avoidance behaviors by identifying risk situations.</td>
</tr>
</tbody>
</table>

**Therapist Support**

The treatment was delivered by two clinical psychologists (BL and JS) with thorough training in exposure-based CBT and experience in treating AF. The participants had continuous contact with an assigned therapist throughout the 10-week treatment period. The role of the therapist was to guide the patient through the treatment and provide feedback on the homework assignments. To progress to the next treatment module, the participants had to complete their weekly home assignments (eg, read the treatment module and conduct the exposure exercises) and report it to the therapist. In addition to reporting the weekly homework assignments, participants could send questions to their therapist via an asynchronous messaging system. Participants received feedback on their reports and questions within two working days. If a participant had technical problems with the platform or did not respond for more than a week, a phone call was made. No treatment interventions were administered over the phone. The treating psychologists could also consult the study cardiologists (FB, HS) at any point throughout the treatment if there were questions regarding the participants’ physical health.

**Procedure**

Cardiologists within tertiary care were informed about the study via email and lectures and referred participants to the study. The referring cardiologists signed a health form where they confirmed fulfillment of the inclusion criteria, meaning that they confirmed the AF diagnosis and the classification of EHRA class ≥IIb; [27]; certified that the participant had undergone a thorough cardiac evaluation including a recent echocardiography (<12 months old; ejection fraction >35%); and ensured that the participants’ medical treatment was in accordance with the current guidelines [3] and that the participants had no contraindications to being physically active. Eligible patients underwent a telephone-based structured psychological assessment by a clinical psychologist (JS) and completed the Alcohol Use Disorders Identification Test [40] and PHQ-9 [35]. Participants then underwent a telephone-based clinical interview by a final-year medical student (SK) who also screened the participants’ medical charts. In case of uncertainty regarding the participants’ physical health, the study cardiologists (FB, HS) were consulted before a decision on inclusion was made, and informed consent was obtained. All self-rated measures were completed over the internet using a secure assessment tool, and the outcomes were collected at pretreatment, posttreatment, and 6-month follow-up. In addition, a shortened version of AFEQT (AFEQT-S; Multimedia Appendix 1), CAQ, and PSS-4 were collected weekly during treatment.

**Statistical Analysis**

All analyses were performed using R [41]. Pretreatment, posttreatment, and 6-month follow-up data were included in a piecewise linear mixed-model analysis with a random intercept. Separate group-level slopes were estimated for the pre- to posttreatment assessment (Slope 1) and posttreatment to the 6-month follow-up assessment (Slope 2). A significant Slope 1 was interpreted as a treatment effect. A nonsignificant Slope 2 was interpreted as the maintenance of improvement during the follow-up period, whereas a significant Slope 2 was
interpreted as improvement or deterioration during the follow-up period. Slope 1 and Slope 2 were then summed to form the estimated overall pre- to 6-month follow-up improvement. Effect sizes (Cohen $d$) were calculated by dividing the estimated pairwise differences between the three time points by the model-implied standard deviation at pretreatment. Effect sizes were categorized according to Cohen recommendations, with small, medium, and large effect sizes corresponding to $d=0.20$, 0.50, and 0.80, respectively [42]. To account for the nonnormal distribution of $d$ [43], 95% CIs for the effect sizes were obtained from 5000 bootstrap replications of the mixed-model analyses.

The ECG measurements were used to analyze the change in AF burden from pretreatment to posttreatment and 6-month follow-up as well as to investigate changes in AF symptom overestimation (ie, indicating AF symptoms on the device when in sinus rhythm). Multimedia Appendix 1 shows a further description of the ECG analysis.

Finally, we performed exploratory mediation analyses to investigate whether changes in symptom preoccupation were a potential mediator of the treatment effect on QoL. We included the three subscales of CAQ (attention, avoidance, and fear) as the indicators of symptom preoccupation, and we also included PSS-4 as a competing mediator. PSS-4 measures stress sensitivity, which was not targeted by the CBT intervention, and by including it that as a competing mediator, we could control for nonspecific improvements. Multimedia Appendix 1 shows for a further description of the mediation analysis.

**Results**

**Sample**

The study comprised 19 participants (12 women and seven men), who were recruited and treated between January 2016 and January 2018. Figure 1 shows the patient flow through the study. The mean age was 60.9 (SD 9.6) years, and the reported mean time since the diagnosis of AF was 6.2 (SD 8.4) years. Table 1 shows the demographics of participants.

During the analysis of the ECG data that were obtained after inclusion, we discovered that two participants had persistent AF at the pretreatment assessment, which means that they, in retrospect, did not fulfill the inclusion criteria of paroxysmal AF at the pretreatment assessment; however, both the participants and their treating physician had reported paroxysmal AF during the inclusion process and were also assessed by the research team as such. Therefore, we decided to keep these two participants in the analysis.

**Figure 1.** Patient flow through the study. AF: atrial fibrillation.
### Table 1. Patient demographics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, n (%)</td>
<td>12 (63)</td>
</tr>
<tr>
<td>Age (in years), mean (SD)</td>
<td>60.9 (9.6)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>7 (37)</td>
</tr>
<tr>
<td>Retired</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Sick leave</td>
<td>3 (16)</td>
</tr>
<tr>
<td>In partnership</td>
<td>9 (47)</td>
</tr>
<tr>
<td><strong>Highest completed education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Vocational</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Post-secondary</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>13 (68)</td>
</tr>
<tr>
<td>AF(a) duration, mean (SD)</td>
<td>6.2 (8.4)</td>
</tr>
<tr>
<td>Previous ablation, n (%)</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Pacemaker, n (%)</td>
<td>2 (11)</td>
</tr>
<tr>
<td><strong>Current medication, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Antiarrhythmics</td>
<td>7 (37)</td>
</tr>
<tr>
<td>Beta blockers</td>
<td>16 (84)</td>
</tr>
<tr>
<td>Anticoagulation</td>
<td>13 (68)</td>
</tr>
<tr>
<td>ACEi(b)/ARB(c)</td>
<td>5 (26)</td>
</tr>
<tr>
<td>Anxiolytics</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Sleep medication</td>
<td>4 (21)</td>
</tr>
<tr>
<td><strong>Medical disorders, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Previous stroke/TIA(d)</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Obstructive sleep apnea</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>5 (26)</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Any medical disease</td>
<td>13 (68)</td>
</tr>
<tr>
<td>Previous psychological treatment</td>
<td>12 (63)</td>
</tr>
<tr>
<td>Current psychiatric disorder</td>
<td>4 (21)</td>
</tr>
</tbody>
</table>

\(a\) AF: atrial fibrillation.
\(b\) ACEi: angiotensin-converting enzyme inhibitor.
\(c\) ARB: angiotensin receptor blocker.
\(d\) TIA: transient ischemic attack.

**Treatment Activity**

In total, 16 out of 19 participants (84%) were considered as treatment completers after completing at least four modules of the treatment and thus commenced work with exposure exercises. The noncompleters (n=3) completed 2-3 modules. The mean number of messages that participants sent and received from their treating psychologist were 12.0 (SD 7.1; range 1-25) and 15.7 (SD 6.0; range 1-27), respectively. The psychologist spent a weekly mean of 9.5 minutes (SD 6.4; range: 4-23 min) per patient.

**Primary and Secondary Outcomes**

Table 2 displays scores on the continuous outcome measures at the assessment points together with \(P\) values of slope estimates and effect sizes (d) with 95% bootstrap CIs. Participants showed significant improvements in AF-specific QoL, as measured by AFEQT, with large within-group effect.
sizes posttreatment \( (d=0.80) \) and at 6-month follow-up \( (d=0.72) \) as compared with baseline, whereas there was no significant difference between posttreatment and 6-month follow-up. The total AFEQT score corresponded to moderate AF severity at baseline; however, at posttreatment and follow-up, the total AFEQT score corresponded to mild AF severity. We observed large effect sizes posttreatment \( (d=1.43) \) and at 6-month follow-up \( (d=1.52) \) in the reduction of cardiac anxiety (CAQ). All three subscales of CAQ (attention, avoidance, and fear) showed a significant improvement posttreatment and at 6-month follow-up. We observed moderately sized pre- to posttreatment improvements in depression (PHQ-9), general anxiety (GAD-7), perceived stress (PSS-4), and general QoL (WHODAS). Furthermore, we observed a medium-sized improvement in self-reported AF symptom severity and a small improvement in AF symptom frequency (SCL scales) and general AF symptoms (AFFS) from pre- to posttreatment. We observed a medium effect size in the reduction of cardiac-specific health care visits posttreatment and a further reduction with a large effect size at 6-month follow-up.
Table 2. Continuous treatment outcome measures.a

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Observed outcomes</th>
<th>Contrasts</th>
<th>Pre to post</th>
<th>Pre to follow-up</th>
<th>Post to follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretreatment, mean (SD)</td>
<td>Posttreatment, mean (SD)</td>
<td>Follow-up, mean (SD)</td>
<td>P value</td>
<td>Cohen d (95% CI)</td>
</tr>
<tr>
<td>AFEQTb</td>
<td>66.8 (18.9)</td>
<td>81.0 (17.4)</td>
<td>82.6 (18.3)</td>
<td>&lt;.001</td>
<td>0.80 (0.45 to 1.24)</td>
</tr>
<tr>
<td>CAQc total</td>
<td>32.5 (10.2)</td>
<td>19.5 (9.2)</td>
<td>18.6 (8.0)</td>
<td>&lt;.001</td>
<td>1.43 (0.91 to 1.94)</td>
</tr>
<tr>
<td>CAQ fear</td>
<td>16.1 (5.8)</td>
<td>9.4 (4.1)</td>
<td>9.3 (3.7)</td>
<td>&lt;.001</td>
<td>1.45 (0.88 to 2.00)</td>
</tr>
<tr>
<td>CAQ avoidance</td>
<td>6.3 (4.3)</td>
<td>3.8 (3.8)</td>
<td>3.1 (2.8)</td>
<td>&lt;.001</td>
<td>0.68 (0.21 to 1.25)</td>
</tr>
<tr>
<td>CAQ attention</td>
<td>10.1 (3.7)</td>
<td>6.3 (3.1)</td>
<td>6.2 (3.4)</td>
<td>&lt;.001</td>
<td>1.12 (0.64 to 1.67)</td>
</tr>
<tr>
<td>SCLf frequency</td>
<td>19.2 (10.9)</td>
<td>15.8 (9.9)</td>
<td>15.2 (10.2)</td>
<td>.013</td>
<td>0.33 (0.09 to 0.78)</td>
</tr>
<tr>
<td>SCL severity</td>
<td>21.2 (6.3)</td>
<td>19.0 (3.6)</td>
<td>19.2 (4.2)</td>
<td>.02</td>
<td>0.46 (0.12 to 0.88)</td>
</tr>
<tr>
<td>AFSSa</td>
<td>10.3 (7.8)</td>
<td>8.1 (6.6)</td>
<td>7.6 (6.5)</td>
<td>.08</td>
<td>0.27 (0.02 to 0.72)</td>
</tr>
<tr>
<td>AFSS health care visits</td>
<td>1.3 (2.0)</td>
<td>0.6 (0.8)</td>
<td>0.2 (0.4)</td>
<td>.046</td>
<td>0.56 (0.12 to 0.92)</td>
</tr>
<tr>
<td>PHQ-9j</td>
<td>6.1 (5.4)</td>
<td>3.1 (3.3)</td>
<td>3.9 (2.8)</td>
<td>.002</td>
<td>0.75 (0.35 to 1.32)</td>
</tr>
<tr>
<td>GAD-7l</td>
<td>6.1 (5.8)</td>
<td>3.4 (3.1)</td>
<td>3.8 (3.3)</td>
<td>.005</td>
<td>0.64 (0.23 to 1.06)</td>
</tr>
<tr>
<td>PSS-4k</td>
<td>5.7 (3.4)</td>
<td>4.1 (2.4)</td>
<td>4.6 (2.2)</td>
<td>.006</td>
<td>0.62 (0.23 to 1.07)</td>
</tr>
<tr>
<td>WHODASl</td>
<td>19.0 (17.0)</td>
<td>12.8 (16.8)</td>
<td>11.0 (14.6)</td>
<td>.004</td>
<td>0.38 (0.12 to 0.86)</td>
</tr>
</tbody>
</table>

aWithin-group effect sizes (ES; Cohen d) and P-values are presented for differences between the three assessment points: pretreatment, posttreatment, and 6-month follow-up based on the piecewise linear mixed-model analysis. ES were reported with 95% CIs based on 5000 bootstrap replications.

bAFEQT: Atrial Fibrillation Effect on Quality-of-Life.

cP < .001.

dCAQ: Cardiac Anxiety Questionnaire.

eP < .01.

fSCL: Symptoms Checklist, Frequency and Severity Scale.

gP < .05.

hAFSS: University of Toronto Atrial Fibrillation Severity Scale.

iPHQ-9: Patient Health Questionnaire-9.

jGAD-7: Generalized Anxiety Disorder-7.

kPSS-4: Perceived Stress Scale-4.

lWHODAS: World Health Organization Disability Assessment Schedule.

**ECG Evaluation**

A total of 2243 ECG recordings were documented. Of these, 2189 (97.6%) could be classified as sinus rhythm or AF, 36 (1.6%) could not be classified, and 18 (0.8%) were considered artifacts. One participant contributed only pretreatment observations. The AF burden increased from pretreatment to the 6-month follow-up (OR=1.235; P=.02). In the AF overestimation analysis, 2175 observations, of which 143 (6.6%) were symptom indications, were included. We excluded 14 recordings where patients were in sinus rhythm but had an irregular pulse due to premature ventricular contractions, which may cause symptoms similar to those of AF, from the analysis. We observed statistically significant decreases in overestimation proportions at both posttreatment (OR=−1.153, P<.001) and 6-month follow-up (OR=−1.538; P<.001), compared with the pretreatment assessment. Multimedia Appendix 1 shows the
observed AF burden and overestimation of AF symptoms at the three assessment points and odds ratios from the multilevel logistic regressions.

Mediation Analysis
Mediators and outcomes were collected for weekly measurements for 9 consecutive weeks during treatment. The mean number of observations per week was 16.9 (out of possible 19), and the lowest number of observations was 15 during the eighth week. In the single mediator analysis, all the tested mediators had statistically significant ab-products, thereby indicating that they all explained some part of the improvement during treatment on the AFEQT-S. In the multiple mediator analysis, where the mediators competed with each other in explaining the change in AFEQT-S, the CAQ fear subscale and PSS-4 ab-products were substantially lower than that in the single mediator analysis, and the former was no longer significant. The ab-products for the CAQ attention and avoidance subscales were of equal size, statistically significant, and more than three times larger than the other two mediators in the multiple mediator model. Multimedia Appendix 1 shows the estimated indirect effects, ab-products, and their confidence intervals for the four mediators, when tested separately and together in a multiple mediator model.

Treatment Satisfaction
At posttreatment, participants scored an average of 24.9 (SD 4.9) out of 32 points on treatment satisfaction as measured by the CSQ-8, and 18 out of 19 participants reported that they were satisfied with the treatment and that it had helped them to deal more effectively with their AF symptoms.

Changes in Medication and in Cardiac Health
At follow-up, 2 of 17 participants had made minor changes in their cardiac medication. None of the patients had undergone invasive cardiac procedures between posttreatment and follow-up. At follow-up, one participant reported deterioration in cardiac health due to more frequent AF episodes. The ECG recording at follow-up indicated a progression to persistent AF in one additional patient.

Adverse Events
At posttreatment, four participants reported an adverse event from participating in the study. The adverse events described were stress because of the study (n=3) and an increased cardiac attention (n=1). Three of the adverse events were given a low severity rating, both when the effect occurred and on residual discomfort, corresponding to a rating of 1 on a scale of 0-3. One event was rated as having the highest severity (3/3) at the time of occurrence and a (2/3) moderate severity rating for the residual comfort. At the six-month follow-up, no current or residual adverse events from participating in the study were reported.

Discussion

Principal Findings
To the best of our knowledge, this is the first study to investigate the efficacy and feasibility of internet-delivered, exposure-based CBT for symptom preoccupation in patients with symptomatic paroxysmal AF. We observed a large improvement in the primary outcome measure of AFEQT that measured AF-specific QoL and self-reported AF symptoms and a large reduction in symptom fear, avoidance behaviors, and hypervigilance as measured by the CAQ. These effects remained large-6 months after the treatment. We also observed small to medium significant effects on the other outcome measures at posttreatment, which were sustained or improved at the 6-month follow-up. Furthermore, the participants’ high adherence to and satisfaction with the treatment, limited reports of adverse events, and perception of being able to deal more effectively with their AF symptoms indicate that exposure-based internet-CBT is a feasible treatment option for the target population.

The results suggest that the AF-specific exposure-based internet-CBT had effects in patients with paroxysmal AF that are comparable with the effects of exposure-based internet-CBT in other somatic functional disorders [38,44-47]. Our results also converge with the results of face-to-face and internet-CBT for patients with cardiovascular disease with anxiety and depression reported in a recent meta-analysis [25] and two recent RCTs [48,49].

We observed a significant increase in the objectively measured AF burden at the 6-month follow-up, which is not unexpected because AF is known to be a progressive disease [50]. Interestingly, the overestimation proportion (ie, when the participant indicated AF symptoms while in normal cardiac rhythm) was significantly decreased both at posttreatment and at 6-month follow-up. Psychological distress has been shown to be predictive of overestimating the AF symptoms [11-13], and it is possible that reduction in symptom fear and hypervigilance improved the patients’ ability to accurately estimate their cardiac rhythm. Despite the observed increase in objective AF burden over time, the AF-specific QoL and symptom preoccupation still showed large improvements and were maintained at the 6-month follow-up. Together with the significant reduction in health care visits, this result indicates that despite the natural progression in AF symptoms, participants did not relapse in symptom preoccupation and were still able to maintain a good QoL.

The mediation analyses showed that a reduction in symptom preoccupation mediated the effect of exposure-based internet-CBT on AF-specific QoL and self-reported AF symptoms. These results are in line with the previous mediation analyses by our group of exposure-based CBT, where we have found that a reduction in avoidance behavior mediates improvement in other chronic health conditions [51-53].

Limitations
The results of this study should be interpreted with the following limitations kept in mind. Importantly, we did not use a control group, which limits the causal inference that can be drawn from the results. Without a control group, we cannot control for the passage of time, the effect of attention from a caregiver, and expectancy of improvement. More than half of the participants had a previous experience of psychological treatment; thus, it is possible that psychological treatment was more suitable for the referred participants than the average patient with AF. The
referral by clinical cardiologists could have contributed to a positive expectancy effect dependent on the caregivers' attitude toward CBT. It is also possible that patients who were generally less prone to respond to treatment were referred to the study, hence contributing to a selection bias. The sample also had disproportionately more women and the subjects were younger than the general AF population [50], further limiting the generalizability of the results. Another limitation is that the handheld ECG is not a continuous measurement of the cardiac rhythm, which may contribute to undetected AF symptoms remaining undetected.

Conclusions
This is the first study in which patients with symptomatic paroxysmal AF were treated with therapist-guided exposure-based internet-CBT. Despite already receiving optimal medical therapy at baseline, we observed large improvements in AF-specific QoL and symptom preoccupation that were sustained at the 6-month follow-up. Our study highlights avoidance behavior and hypervigilance as potentially important psychological factors contributing to symptom severity and low QoL in patients with AF. However, the psychological aspects of AF appear to be underrecognized in the current medical literature and clinical practice.

We conclude that symptom preoccupation is an important target for treatment and that the treatment may be both feasible and clinically effective for the target population. AF-specific CBT delivered via the internet has the potential to reduce health care utilization and improve the well-being of a large group of patients, when current treatment methods have limited effectiveness. These results need to be confirmed in RCTs.

Acknowledgments
This study was supported by grants from the Swedish Research Council (2016–013792016-01379), Region Stockholm (ALF project), and Karolinska University Hospital. None of the funding bodies had any influence on the study design, implementation, data analysis, or interpretation. The authors would like to thank Professor Matteo Bottai for statistical consultation and the participants in the study.

Conflicts of Interest
FB is a study committee member of Biotronik, Medtronic and has received lecture fee from Pfizer, Novartis, St Jude Medical, Orion, and Boehringer. BL is a shareholder of DahliaQomit AB, a company specializing in online psychiatric symptom assessment, and Hedman-Lagerlöf och Ljötsson psykologi AB, a company that licenses cognitive behavior therapy manuals. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Description of electrocardiogram and mediation analysis.
[DOCX File, 21 KB - cardio_v5i1e24524_app1.docx]  

References


35. 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 [FREE Full text] [Medline: 11556941]


Abbreviations

AF: atrial fibrillation

AFEQT: Atrial Fibrillation Effect on Quality-of-Life
CAQ: Cardiac Anxiety Questionnaire
CBT: cognitive behavior therapy
CSQ-8: Client Satisfaction Questionnaire
ECG: electrocardiogram
EHRA: European Heart Rhythm Association
GAD-7: Generalized Anxiety Disorder
PHQ-9: Patient Health Questionnaire
PSS-4: Perceived Stress Scale
QoL: quality of life
RCT: randomized controlled trial
SCL: Symptoms Checklist, Frequency and Severity Scale
WHODAS: World Health Organization Disability Assessment Schedule

©Josefin Särnholm, Helga Skúladóttir, Christian Rück, Sofia Klavebäck, Eva Ólafsdóttir, Susanne S Pedersen, Frieder Braunschweig, Brjánn Ljóttsson. Originally published in JMIR Cardio (http://cardio.jmir.org), 02.03.2021. 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 http://cardio.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: The internet has become a popular platform for patients to obtain information and to review the health care providers they interact with. However, little is known about the digital footprint of vascular surgeons and their interactions with patients on social media.

Objective: This study aims to understand the activity of academic vascular surgeons on physician rating websites.

Methods: Information on attending vascular surgeons affiliated with vascular residency or with fellowships in the Southern Association for Vascular Surgery (SAVS) was collected from public sources. A listing of websites containing physician ratings was obtained via literature reviews and Google search. Open access websites with either qualitative or quantitative evaluations of vascular surgeons were included. Closed access websites were excluded. Ranking scores from each website were converted to a standard 5-point scale for comparison.

Results: A total of 6238 quantitative and 967 qualitative reviews were written for 287 physicians (236 males, 82.2%) across 16 websites that met the inclusion criteria out of the 62 websites screened. The surgeons affiliated with the integrated vascular residency and vascular fellowship programs in SAVS had a median of 8 (IQR 7-10) profiles across 16 websites, with only 1 surgeon having no web presence in any of the websites. The median number of quantitative ratings for each physician was 17 (IQR 6-34, range 1-137) and the median number of narrative reviews was 3 (IQR 2-6, range 1-28). Vitals, WebMD, and Healthgrades were the only 3 websites where over a quarter of the physicians were rated, and those rated had more than 5 ratings on average. The median score for the quantitative reviews was 4.4 (IQR 4.0-4.9). Most narrative reviews (758/967, 78.4%) were positive, but 20.2% (195/967) were considered negative; only 1.4% (14/967) were considered equivocal. No statistical difference was found in the number of quantitative reviews or in the overall average score in the physician ratings between physicians with social media profiles and those without social media profiles (departmental social media profile: median 23 vs 15, respectively, \( P = .22 \); personal social media profile: median 19 vs 14, respectively, \( P = .08 \)).

Conclusions: The representation of vascular surgeons on physician rating websites is varied, with the majority of the vascular surgeons represented only in half of the physician rating websites The number of quantitative and qualitative reviews for academic vascular surgeons is low. No vascular surgeon responded to any of the reviews. The activity of vascular surgeons in this area of social media is low and reflects only a small digital footprint that patients can reach and review.
KEYWORDS
internet; patient satisfaction; quality of care; physician rating sites; patient experience; professional reviews; social media

Introduction

Social media resources continue to be used by commercial and nonprofit organizations for social marketing because of the advantages of access to a large number of consumers, presence of transparency, the wide global reach, the ability to boost web page traffic, and the ability to promote the brand name [1]. With reference to health care, social media tools play an essential role in reputation management, public outreach, health promotion, and patient education [2]. An additional tool is the provision of quantitative and qualitative ratings of physician performance on physician rating websites that can be shared publicly. Seventy percent of the top 10 Google search results on specific physicians are third-party websites such as physician rating websites [3].

Physician rating websites display valuable information for a consumer regarding a physician's practice, including the physician’s area of expertise, office location, office hours, insurance accepted, in addition to quantitative and qualitative reviews from past consumers. Although surveys have indicated that the insurance accepted, referrals from primary care physicians, and reputation are important in the selection of physicians, 65% of the physician rating website users have been shown to choose a physician after viewing positive reviews on websites, and conversely, 52% of the physician rating website users have avoided providers owing to the negative reviews shown on the websites [4,5].

Although physician rating websites contain a broad range of physician-specific data, the physicians rated on these websites are not evenly distributed among medical specialties [6-16]. While orthopedic surgeons are well represented on physician rating websites with over 90% of the surgeons having ratings on the most popular physician rating websites, radiologists, who have limited direct face-to-face contact with patients, have a very small digital physician rating website footprint with only 20% rated on any of the 5 physician rating websites studied [7-9,11,13]. Multiple studies have demonstrated differences in the ratings across multiple specialties. Physicians in the fields of cardiac surgery, nephrology, genetics, and radiology receive higher ratings, whereas those in addiction medicine, dermatology, neurology, and psychiatry receive lower ratings [16]. Vascular surgery is emerging as a distinct subspecialty and is currently undergoing a branding and identity campaign. In light of this change in the perception of this specialty, there exists a knowledge gap on the digital footprint and performance of vascular surgeons across the spectrum of physician rating websites. Understanding this gap can offer this specialty a roadmap to improve public perception and will prompt further research on the effect of branding and marketing on the ratings within the physician rating websites for vascular surgery. The aim of this study was to examine the accuracy of professional demographics, the presence and responsiveness of academic vascular surgeons across open access physician rating websites, and the quantity and quality of patient reviews within a defined geographic region. The aim was also to define the digital physician rating website footprint of vascular surgery to ascertain whether academic vascular surgeons have evolved to embrace and participate in these reviews.

Methods

This is a cross-sectional study of publicly held data in search engines and websites accessed from a US internet service provider from September 2019 to November 2019. Websites containing physician ratings were obtained via literature reviews and Google search. This study examines data in the public domain and does not contain Health Insurance Portability and Accountability Act (HIPAA) information or interactions with individuals; thus, it is exempt from institutional review board approval or consent.

The current integrated vascular residency and vascular fellowship program lists were obtained from the Accreditation Council for Graduate Medical Education website. Attending vascular surgeons affiliated with each of these programs in the Southern Association for Vascular Surgery (SAVS) were collected [17]. Websites reporting physician ratings were obtained via literature reviews and Google search by using the terms “rate doctors,” “MD review,” “physician ranking,” “doctor rating,” “find doctors,” and “best doctors.” The inclusion criteria were as follows: (1) open access websites and (2) websites that allowed qualitative or quantitative reviews by patients. The exclusion criteria were as follows: (1) physician rating websites outside of the United States; (2) websites limited to a certain geographic area in the United States; (3) websites that excluded vascular surgeons; (4) websites linked to a health care system, and (5) websites that were inaccessible.

Health care review websites obtain the physician list through public records from the National Provider Identifier Registry, medical boards, etc. General review websites allow users or business owners to add physicians or edit information, whereas health care review websites require changes to be made through the management team. Most websites allow physicians to claim their profile for free after they create an account and go through the prompted steps. These websites then allow physicians to manage their profiles, audit for accuracy, respond to reviews, and dispute reviews depending on the website. Some websites offer sponsored profiles to physicians for a fee to promote their practice. Multimedia Appendix 1 provides the definitions of the different profiles of the physicians.

Data collection was performed between August 2019 and September 2019. Physician search was performed on individual websites with physician first name/first initial, middle name/middle initial, last name, and location in various combinations and orders. Supplementary Google search was performed with "physician name, website domain name" to enhance the discovery of the physician's profile. A similar strategy was used in finding physician or department profiles on social media websites such as Facebook, Twitter, and LinkedIn. Physician-specific information such as gender, age, educational background, training, and professional association

https://cardio.jmir.org/2021/1/e22975
was collected through their professional websites affiliated with their institutions and physician rating websites.

Most physician rating websites used a ranking score of 1-5 while Dr.Score and Healthcare reviews used a ranking score of 1-10; the scores of these 2 sites were converted to the standard 5-point scale (1-5) for comparison. Physician rating websites with more than 1000 quantitative reviews were used to examine correlation. The Kendall rank correlation coefficient was used to adjust for ties. USNews&World Report was excluded from this analysis because the ratings in this website were derived from over 100 unspecified web-based sources gathered by a private company and does not collect patient reviews or ratings directly. Descriptive data were presented as median (IQR). Kendall correlation coefficient was used to evaluate the correlation between continuous variables. Mann-Whitney U or Kruskal-Wallis test with post hoc Wilcoxon test was used for continuous data. Chi-square test was used to compare categorical data with a Bonferroni adjustment used for all post hoc tests with adjusted P values reported. Data analysis was performed in RStudio version 1.2.5001 (RStudio Inc).

Results

Review of the Commercial Websites

Sixteen out of 62 websites met the inclusion criteria (Multimedia Appendix 2). Figure 1 shows the flow diagram of the websites included and excluded in this study. WebMD and USNews&World Report were the 2 sites that listed the most number of physicians (271/287, 94.4%) while Yellowbot (48/287, 16.7%) had the least number of physicians (Table 1). Websites had specific search strategies that provided higher yield; however, none of the websites provided instructions. Table 2 shows the search strategy on the 4 most common physician rating websites. Sixteen physician rating websites were included after screening 62 initial websites.

Figure 1. Flow diagram for website inclusion.
Table 1. Physician profiles across 16 physician rating websites (N=287).

<table>
<thead>
<tr>
<th>Websites</th>
<th>Physicians with profile, n (%)</th>
<th>Incomplete profiles, n (%)</th>
<th>Inaccurate profiles, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebMD</td>
<td>271 (94.4)</td>
<td>25 (9.6)</td>
<td>27 (10.0)</td>
</tr>
<tr>
<td>USNews&amp;World Report</td>
<td>271 (94.4)</td>
<td>1 (0.4)</td>
<td>11 (4.1)</td>
</tr>
<tr>
<td>Caredash</td>
<td>266 (92.7)</td>
<td>7 (2.6)</td>
<td>21 (7.9)</td>
</tr>
<tr>
<td>Vitals</td>
<td>264 (92.0)</td>
<td>3 (1.1)</td>
<td>19 (7.2)</td>
</tr>
<tr>
<td>Healthgrades</td>
<td>264 (92.0)</td>
<td>24 (9.1)</td>
<td>44 (16.7)</td>
</tr>
<tr>
<td>YP.com</td>
<td>209 (72.8)</td>
<td>86 (41.1)</td>
<td>23 (11.0)</td>
</tr>
<tr>
<td>RateMD</td>
<td>199 (69.3)</td>
<td>156 (78.4)</td>
<td>50 (25.1)</td>
</tr>
<tr>
<td>Dr.Score</td>
<td>147 (51.2)</td>
<td>46 (31.3)</td>
<td>28 (19.0)</td>
</tr>
<tr>
<td>Insiderpages</td>
<td>101 (35.2)</td>
<td>101 (100.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Local</td>
<td>90 (31.4)</td>
<td>87 (96.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Zocdoc</td>
<td>79 (27.5)</td>
<td>11 (13.9)</td>
<td>4 (5.1)</td>
</tr>
<tr>
<td>Yellowbook</td>
<td>74 (25.8)</td>
<td>72 (97.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Healthcare reviews</td>
<td>70 (24.4)</td>
<td>22 (31.4)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Wellness</td>
<td>62 (21.6)</td>
<td>26 (41.9)</td>
<td>17 (27.4)</td>
</tr>
<tr>
<td>Yelp</td>
<td>53 (18.5)</td>
<td>53 (100.0)</td>
<td>10 (18.9)</td>
</tr>
<tr>
<td>Yellowbot</td>
<td>48 (16.7)</td>
<td>46 (95.8)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

aTotal number of physicians with profiles on the 16 websites=286.

bProfiles (n=766) with less than 50% of the required information (eg, training, expertise) or profiles that lack any physician-specific information besides practice location and office contact information. Percentage was calculated as the number of physicians with incomplete profiles by the total number of physicians with profiles for each website.

cProfiles with any inaccuracy in physician practice and demographic information (n=252). Percentage was calculated as the number of physicians with inaccurate profiles by the total number of physicians with profiles for each website.

Table 2. Search strategy on the 4 most common physician rating websites (N=287).

<table>
<thead>
<tr>
<th>Website name</th>
<th>Name search</th>
<th>Location search</th>
<th>Required search on Google, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vital</td>
<td>Must use middle initial, does not have full middle name in the system. If first name in the system is also initial only, entering full name will not find the physician.</td>
<td>Has a certain degree of matching by name during the search regardless of the correct location.</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Healthgrades</td>
<td>Name must have the exact match being “first name, last name” in order to find through search. However, will provide matching in the search box without such restriction.</td>
<td>Has a certain degree of matching by name in the search box regardless of the correct location.</td>
<td>7 (2.4)</td>
</tr>
<tr>
<td>WebMD</td>
<td>Good name match. No requirement for the order of the name or differentiation between initial versus expanded name.</td>
<td>Location needs to be correct. Very low degree of matching by name only.</td>
<td>18 (6.3)</td>
</tr>
<tr>
<td>USNews&amp;World Report</td>
<td>Must have correct expanded first name and last name. Order of the name and the middle name does not affect search.</td>
<td>Match of name is sufficient, correct location not required.</td>
<td>1 (0.3)</td>
</tr>
</tbody>
</table>

Review of the Academic Divisional Participation
One potential confounding factor in individual physician profiles is the corresponding activity of the divisional profile digital footprint. Nine (9/37, 24%) vascular surgery divisions had social media profiles: 7 had only Twitter accounts, 1 had only Facebook profiles, and 1 had both Twitter and Facebook profiles. Physicians in the institutions with established departmental social media websites had a higher number of ratings compared to those without established departmental social media websites (median 23 [IQR 5-38] vs 15 [IQR 7-33], respectively, $P=.22$), although this was not statistically significant.

Review of Individual Physician Participation
Surgeons affiliated with the integrated vascular residency and vascular fellowship program of SAVS had a median (IQR) of 8 (7-10) profiles across 16 websites with only 1 surgeon having
no web presence on any of the sites. Most physician profiles (2214/2466, 89.8%) were accurate, reflecting correct demographic affiliations and practice information (Table 1). Of the 8 websites wherein claimed profiles were clearly distinguished, only 14.9% (43/287) of the physicians considered in this study had at least one claimed profile. Of the single website that had clear notations of the sponsored profiles on the profile page itself, no vascular surgeon had sponsored profiles. Of the 287 physicians, 115 (40.1%) were members of the Society of Vascular Surgery and 82 (28.6%) were members of other vascular societies, while the remainder did not disclose their affiliations; 195 (67.9%) physicians had profiles on at least one social media platform: 57.5% (165/287) on LinkedIn, 19.9% (57/287) on Twitter, and 18.5% (53/287) on Facebook.

A total of 6238 ratings and 967 narrative reviews were written for 287 physicians (236 males, 82.2%) affiliated with the integrated vascular residency and vascular fellowship program within the SAVS across the 16 websites surveyed (Table 3 and Table 4). The median number of quantitative ratings for each physician among those with at least 1 rating was 17 (IQR 6-34, range 1-137) and the median number of narrative reviews among those with at least 1 narrative review was 3 (IQR 2-6, range 1-28); 12.9% (37/287) of the physicians had 0 quantitative reviews and 31.0% (89/287) had 0 qualitative reviews. Ratings were overwhelmingly positive, with a median weighted average score of 4.4 (IQR 4.0-4.9) out of a total score of 5. Most narrative reviews (758/967, 78.4%) were also positive, but 20.2% (195/967) of them were considered negative; only 1.4% (14/967) were considered equivocal. Physicians with negative narrative reviews had lower ratings compared to those without (median 4.07 vs 4.7, \( P = .001 \)). There was no physician response to any patient review.

Table 3. Physician rating scores across 16 physician rating websites (N=287).

<table>
<thead>
<tr>
<th>Websites</th>
<th>Physicians with rating, n (%)a</th>
<th>Total quantitative ratings (n=6238), n</th>
<th>Median score</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitals</td>
<td>217 (75.7)</td>
<td>1731</td>
<td>4.4</td>
<td>4-5</td>
</tr>
<tr>
<td>WebMD</td>
<td>200 (69.7)</td>
<td>1737</td>
<td>4.5</td>
<td>4-5</td>
</tr>
<tr>
<td>Healthgrades</td>
<td>193 (67.2)</td>
<td>1193</td>
<td>4.6</td>
<td>3-7-5</td>
</tr>
<tr>
<td>USNews&amp;World Report</td>
<td>71 (24.7)</td>
<td>1296</td>
<td>5</td>
<td>4-5</td>
</tr>
<tr>
<td>RateMD</td>
<td>46 (16.0)</td>
<td>103</td>
<td>4</td>
<td>3-5</td>
</tr>
<tr>
<td>Dr.Score</td>
<td>24 (8.3)</td>
<td>34</td>
<td>5</td>
<td>4-5</td>
</tr>
<tr>
<td>Caredash</td>
<td>18 (6.3)</td>
<td>22</td>
<td>5</td>
<td>5-5</td>
</tr>
<tr>
<td>Wellness</td>
<td>11 (3.8)</td>
<td>20</td>
<td>4.5</td>
<td>3-8-5</td>
</tr>
<tr>
<td>Yellowbot</td>
<td>7 (2.4)</td>
<td>17</td>
<td>5</td>
<td>5-5</td>
</tr>
<tr>
<td>Zocdoc</td>
<td>7 (2.4)</td>
<td>65</td>
<td>5</td>
<td>4-5</td>
</tr>
<tr>
<td>Insiderpages</td>
<td>6 (2.1)</td>
<td>13</td>
<td>5</td>
<td>4-3-5</td>
</tr>
<tr>
<td>YP.com</td>
<td>4 (1.4)</td>
<td>4</td>
<td>5</td>
<td>4-5</td>
</tr>
<tr>
<td>Yelp</td>
<td>2 (0.7)</td>
<td>2</td>
<td>3</td>
<td>2-4</td>
</tr>
<tr>
<td>Healthcare reviews</td>
<td>1 (0.3)</td>
<td>1</td>
<td>5</td>
<td>N/Ab</td>
</tr>
<tr>
<td>Local</td>
<td>0 (0.0)</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yellowbook</td>
<td>0 (0.0)</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aTotal number of physicians with at least 1 rating=251.

bN/A: not applicable.
Table 4. Narrative reviews across 16 physician rating websites.

<table>
<thead>
<tr>
<th>Websites</th>
<th>Physicians with narrative reviews (n=287), n (%)a</th>
<th>Total narrative reviews (n=967), n</th>
<th>Positive narrative reviews, n (%)b,c</th>
<th>Neutral narrative reviews, n (%)d,e</th>
<th>Negative narrative reviews, n (%)e</th>
<th>Patient found useful (n=737), n</th>
<th>Patient found not useful (n=150), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitals</td>
<td>151 (52.6)</td>
<td>482</td>
<td>374 (77.6)</td>
<td>5 (1.0)</td>
<td>103 (21.4)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WebMD</td>
<td>1 (0.3)</td>
<td>1</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Healthgrades</td>
<td>136 (47.4)</td>
<td>308</td>
<td>248 (80.5)</td>
<td>0 (0.0)</td>
<td>60 (19.5)</td>
<td>704</td>
<td>148</td>
</tr>
<tr>
<td>USNews&amp;World Report</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RateMD</td>
<td>43 (15)</td>
<td>88</td>
<td>58 (65.9)</td>
<td>5 (5.7)</td>
<td>25 (28.4)</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Dr.Score</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Caredash</td>
<td>14 (4.9)</td>
<td>16</td>
<td>15 (93.8)</td>
<td>1 (6.3)</td>
<td>0 (0.0)</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Wellness</td>
<td>8 (2.8)</td>
<td>17</td>
<td>15 (88.2)</td>
<td>0 (0.0)</td>
<td>2 (11.8)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yellowbot</td>
<td>7 (2.4)</td>
<td>19</td>
<td>19 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zocdoc</td>
<td>6 (2.1)</td>
<td>29</td>
<td>23 (79.3)</td>
<td>3 (10.3)</td>
<td>3 (0.3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Insiderpages</td>
<td>1 (0.3)</td>
<td>1</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>YP.com</td>
<td>4 (1.4)</td>
<td>4</td>
<td>3 (75.0)</td>
<td>0 (0.0)</td>
<td>1 (25.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yelp</td>
<td>2 (0.7)</td>
<td>2</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>1 (50.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Health care reviews</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Local</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yellowbook</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a197 physicians received narrative reviews.
bTotal number of positive narrative interviews=758.
cThe percentages for the positive, neutral, and negative narrative review columns were calculated over the total narrative reviews in the 3rd column.
dTotal number of neutral narrative interviews=14.
eTotal number of negative narrative interviews=195.

Correlation Between the Rating Scores

The physician scores on Vitals and WebMD correlated well (Kendall τ=0.78, P<.001) while those on Healthgrades correlated poorly with Vitals (Kendall τ=0.15, P=.007) and WebMD (Kendall τ=0.17, P=.006). Years of experience (Kendall τ =0.12, P=.007), personal social media profile (Kendall τ=0.03, P=.57), departmental social media profile (Kendall τ=−0.05, P=.34), and number of ratings (Kendall τ=−0.14, P=.001) did not correlate or they only weakly correlated with the weighted average score.

Physician and Practice Factors Affecting the Size of the Digital Footprint

The social media profiles of vascular surgeons (especially LinkedIn) and the rating number followed a similar curve with small peaks in the age groups of 20-24 years and 35-39 years (Figure 2 and Figure 3). Physicians with individual social media profiles had a higher median number of ratings compared to those who did not; however, this did not reach statistical significance (19 vs 14, respectively, P=.08). Fewer female vascular surgeons had personal social media profiles (P=.02). Female surgeons had fewer years of experience (median 14 vs 24 for males, respectively, P=.001), fewer profiles (7.5 vs 9, respectively, P=.02), fewer number of ratings (median 6 vs 19, respectively, P<.001) but a similar number of narrative reviews (3 vs 4, respectively, P=.23). Physicians in a practice with 10 vascular surgeons or more had more ratings (21 vs 13, respectively, P=.01) but similar number of profiles (median 9 vs 8, respectively, P=.09) and narrative reviews (4 vs 3, respectively, P=.18).

https://cardio.jmir.org/2021/1/e22975

JMIR Cardio 2021 | vol. 5 | iss. 1 | e22975 | p.84

XSL•FO
RenderX
**Figure 2.** Total number of ratings across 16 physician rating websites by years of experience (number of years after graduation from allopathic or osteopathic school). The lower, middle, and upper hinges of the box represent 25th percentile, 50th percentile or median, and 75th percentile, respectively. Whiskers represent 1.5 interquartile range, and points represent outliers.

**Figure 3.** Proportion of physicians with profiles on 3 different social media platforms among physicians with different years of experience by the number of years after graduation from an allopathic or osteopathic school.
Reviewer Information

Of the 2 physician rating websites with a significant number of narrative reviews, Vitals did not provide reviewer names while a third (111/308, 36.0%) of the reviewers on Healthgrades provided their full name; another 29.5% (91/308) provided only the initial or the first name, while a third (106/308, 34.4%) remained completely anonymous.

Discussion

Principal Results

Our study found that the presence of academic vascular surgeons affiliated with the integrated vascular residency and vascular fellowship program within the SAVS was low on physician rating websites. Of the 287 physicians, 87.4% (251/287) of the physicians had at least one rating across 16 physician rating websites and 68.6% (197/287) had at least one narrative review. The density of the reviews was low, with a median of 17 for the quantitative ratings and a median of 3 for the narrative reviews for those who had at least 1 rating or narrative review across 16 physician rating websites. No physician responded to any of the reviews.

Limitations of This Study

The study has a limited sample and may not be extrapolated to all US academic vascular physicians. We were unable to correlate physician ratings with USNews&World Report hospital rankings, urban or rural locations, and whether the academic center had a social media department because minimal variation existed in these variables. It was not possible to identify the type of procedure offered by the physicians to assess its effect on the patient reviews. We did not correlate hospitals’ own review systems with physician rating websites or Press Ganey scores. Additionally, information on the internet is dynamic; physicians could have accumulated additional reviews since the time of our data collection. Lastly, we did an extensive search to investigate the physician rating websites, and multiple searches were performed on individual providers to ensure that we captured maximum information. However, it is possible that there were websites or profiles that we may have missed or we may not have investigated. This likely does not significantly affect our study, as it would be equally difficult for patients to locate such websites or profiles and likely contain minimal information.

Comparison With Prior Work

Direct comparisons between previous studies of physician rating websites is difficult due to differences in physician selection. There is a general trend of an increasing percentage of physicians rated over time. Vascular surgeons were significantly underrepresented compared to orthopedic surgeons; 94.3%-99.5% of the orthopedic surgeons were rated on physician rating websites with a much higher number of ratings—triple that of vascular surgeons in some instances, and we speculate that this difference may be attributed to differences in the specialties rather than differences in physician demographics [7-9,11,12,18]. The cluster of studies on the digital footprint of orthopedic surgeons reflects better awareness of social media, more effort in the web-based promotion of their practice, and thus, a larger digital footprint [7-9,11,12,18,19].

In general, patients rated vascular surgeons very positively, with the median score for almost all the websites being 4 or higher on a 5-point scale, and 78.4% (758/967) of the comments were positive overall. This is consistent with prior findings regardless of specialty [6,11,20-22]. Despite the overall positivity of the reviews, the consistency of the ratings at the physician level across physician rating websites was variable. Similar to that reported in previous studies, Vitals and WebMD had excellent correlation, but both correlated poorly with Healthgrades [8,11].

In our study, we did not find any physician demographic characteristics, level of social media presence, or total rating frequency that contributed to the overall rating scores. Some studies found that younger physicians received higher scores, which could be attributed to the better relationships between younger patients and younger physicians, thereby leading to increased number of reviews with high scores [8,11,23]. The associations among gender, total rating frequency, online presence, and rating score varied among studies [6,8,9,11,15,20,21,23-25].

Higher numbers of physician rating website profiles were seen for vascular surgeons with social media profiles, whether personal or departmental, but there was no statistical significance. Physicians with <14 or >40 years of experience were less likely to be rated. Small peaks were noted in the 20-24 years and 35-39 years of experience groups. This could be related to the similar trend seen in social media profiles. Gao et al also found that less experienced physicians had fewer ratings because they are still developing their practice and reputation [23]. Female vascular surgeons had fewer social media profiles and fewer physician rating website ratings compared to their male counterparts, which may be due to the younger age, and this finding has been reported in previous studies [11,20]. We found that physicians in practices with 10 or more surgeons had more ratings, but this finding has not been reported in previous studies. This could be related to large practices located in densely populated areas, which have higher number of ratings compared to the less densely populated regions reported in a recent study [16]. We did not examine metropolitan versus nonmetropolitan locations of practices ourselves because most of the academic affiliated practices were located in areas considered to be metropolitan.

The inconsistency in the physician ratings between physician rating websites and variable findings in factors predictive of better ratings is likely due to the low density of the ratings, leading to high susceptibility to outliers. Healthgrades, Vitals, WebMD, USNews&World Report, and Zocdoc were the only sites wherein the average number of ratings exceeded 5 for those rated. Vitals, WebMD, and Healthgrades were the only 3 websites, wherein over a quarter of the physicians were rated. These may be the better websites for vascular surgeons to focus on in a social media campaign. The large number of physician rating websites and directories—33 in 2010 and 28 in 2018—dilutes patient reviews [26,27]. Less than 5% of the physicians were rated on 56% (9/16) of the websites in our study. Less than 1% of the physicians were rated on 82% of the
websites in the study of Lagu et al, and most of these sites were no longer accessible at the time of this study [26]. This rationalization and consolidation in the physician rating website marketplace will lead to an improved density of reviews.

Responding to a review is an important customer relationship exercise. No physician responded to any qualitative review in our study, including the 193 negative narrative reviews. In a study of the German physician rating website called “Jameda,” 1.58% of all the numeric ratings received responses while almost a third of the narrative reviews received responses from a physician [28]. Those physicians who responded to reviews had more “likes” and visits to the Jameda and had better ratings. Although these physicians were also more active on Jameda, improved rating and site travel may not be attributable to responding to patient reviews alone but they confirmed that active participation on physician rating websites has positive effects. Studies on response to patient reviews are limited in the medical field, but multiple studies in the hotel industry have found that responses to negative reviews can mitigate adverse effects [29-31].

Limited physician responses may be related to fear of HIPAA violation and offering an asynchronous medical opinion. Revealing patient information on social media without patient consent undermines patient trust and can lead to legal and civil actions [32]. However, most negative reviews are organizational issues outside the control of the providers, such as wait time, accessibility, and difficulty making appointment [10,26,33]. The best practices to actively respond to negative feedback is to respond offline or speak in the web-based platform in general terms, avoid confirming or denying the person as a patient, acknowledge the complaint issue, apologize, and provide an action plan [34,35]. Appropriate response to an active negative review can gain the trust of prospective patients auditing previous patient comments and reviews. In addition, the American Medical Association, along with others, recommend politely asking patients for reviews to dilute negative reviews since the majority will be positive [6,36].

Narrative reviews on physician rating websites have the potential of expanding the scope of quality measurement for providers. A recent study of Yelp reviews on hospitals has shown that patient reviews not only covered 7 out of the 12 categories included in the Hospital Consumer Assessment of Healthcare Providers and Systems but also consisted of additional 12 categories (Textbox 1) [37]. Studies have shown that physicians have used patient web-based reviews to improve patient care in areas of patient communication, scheduling process, and office workflow [38,39]. Narrative reviews on physician rating websites have the same potential of expanding the scope of quality measurements and aiding in quality improvement.

The limitations of physician rating websites include varied quality, limited accountability, and limited representativeness. In a global study on physician rating websites, the United States had a large number of physician rating websites but more considerable variation in the quality compared to European physician rating websites [40]. Anonymity, while beneficial for allowing uncensored speech, results in decreased accountability and false reviews such as negative reviews from competitors or self-written positive reviews [26]. While a third of patients provided their full name on Healthgrades, some appeared to be fake. This anonymity makes it difficult to validate their reviews or complaints on physician rating websites. The difficulty in validating and managing public responses is likely a reason that many hospitals do not engage in patient reviews. Furthermore, the web-based reviewer is not a random sample of the patient pool but rather an impressed or aggrieved party. There also exists demographic characteristics associated with physician-rating behavior and gender bias in perceptions of patient-physician interactions [25,41]. However, physician rating websites in the United States lack basic patient/reviewer demographics to allow further studies. Lastly, physician rating websites need to improve transparency by disclosing what enhancements are provided for sponsored profiles and clearly distinguish sponsored profiles not only during unspecific searches but also in specific searches and document it on the individual profiles.

Social media platforms have been used to advertise practice and attract referrals with relatively high returns on investment [42]. The physician rating website is an important part of these social media platforms. Third-party physician information websites, including the physician rating website, make up the majority of the top search results on Google [43]. Physician rating websites can attract patients by showing prior patient experiences and opinions, which can be exceedingly important in elective settings. Additionally, in concert with a well-designed social media strategy, physician rating websites can be a great platform to showcase a wide range of procedures that vascular surgeons perform alone or as assistance for other specialties. Given the findings in this study and a review of the current literature, we recommend the following physician rating website management strategies to improve value proposition: (1) focus on a limited number of physician rating websites (Vitals, Healthgrades, and WebMD), (2) assume the management profiles on physician rating websites to ensure the accuracy of information and allow physicians to receive instantaneous feedback, (3) invite patients to write reviews on these websites, and (4) develop a response strategy to reviews on the physician rating website chosen.
Textbox 1. Yelp domains that can supplement and inform traditional surveys of the patient experience of care.

<table>
<thead>
<tr>
<th>Yelp domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Cost of hospital visit</td>
</tr>
<tr>
<td>• Insurance billing</td>
</tr>
<tr>
<td>• Ancillary testing</td>
</tr>
<tr>
<td>• Facilities</td>
</tr>
<tr>
<td>• Amenities</td>
</tr>
<tr>
<td>• Scheduling</td>
</tr>
<tr>
<td>• Compassion of staff</td>
</tr>
<tr>
<td>• Family member care</td>
</tr>
<tr>
<td>• Quality of nursing</td>
</tr>
<tr>
<td>• Quality of staff</td>
</tr>
<tr>
<td>• Quality of technical aspects of care</td>
</tr>
<tr>
<td>• Specific type of medical care</td>
</tr>
</tbody>
</table>

Conclusion

The representation of vascular surgeons on physician rating websites is varied, with the majority of the vascular surgeons represented only in half of the physician rating websites. The number of quantitative and qualitative reviews for vascular surgeons is low; therefore, no surgeon responded to any of the reviews. The activity of the vascular surgeons in this area of social media is low and reflects a small digital footprint that patients can reach and review. Healthgrades, Vitals, and WebMD are the most recommended physician rating platforms for vascular surgeons to focus on to promote and improve their practice.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of physician profiles.
[DOCX File, 21 KB - cardio_v5i1e22975_app1.docx]

Multimedia Appendix 2
List of physician rating websites included in our study.
[DOCX File, 22 KB - cardio_v5i1e22975_app2.docx]

References

1. Sajid SI. Social Media and Its Role in Marketing. Bus Eco J 2015;07(01) [FREE Full text] [doi: 10.4172/2151-6219.1000203]


https://cardio.jmir.org/2021/1/e22975


Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
SAVS: Southern Association for Vascular Surgery

©Qi Yan, Katherine J Jensen, Rose Thomas, Alyssa R Field, Zheng Jiang, Christian Goei, Mark G Davies. Originally published in JMIR Cardio (http://cardio.jmir.org), 24.02.2021. 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 http://cardio.jmir.org, as well as this copyright and license information must be included.