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; Francoise 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 data collection. The same device was used for all participants and was cleaned between use with a hospital-grade disinfectant. 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.
**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]:

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

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 (\( P < 0.05 \) and further confirmed the average difference was related to the magnitude. We thus calculated the regression-based 95% limits of agreement as \( 27.7922 - 0.3332A - 1.96 \times \text{SD} \) (of the residuals; lower limit) and \( 27.7922 - 0.3332A + 1.96 \times \text{SD} \) (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).
Figure 2. Scatter plots of standard deviation of measurement pair differences against patient mean. (Patients 23, 18, 12 had rapid ventricular response.) Here, we show a relationship between difference and magnitude of measurement, suggesting a violation of constant bias assumption.

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/m$^2$), 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 Schafer Memorial Fund. SM reports personal fees for serving on scientific advisory boards for Akcea Therapeutics, Amgen, and...
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

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by J Raja, M Doerr, A Teh; comments to author 04.03.20; revised version received 30.06.20; accepted 15.12.20; published 08.02.21

Please cite as:

Huynh P, Shan R, Osuji N, Ding J, Isakadze N, Marvel FA, Sharma G, Martin SS
Heart Rate Measurements in Patients with Obstructive Sleep Apnea and Atrial Fibrillation: Prospective Pilot Study Assessing Apple Watch’s Agreement With Telemetry Data
JMIR Cardiol 2021;5(1):e18050
URL: http://cardio.jmir.org/2021/1/e18050/
doi: 10.2196/18050
PMID: 33555260

©Pauline Huynh, Rongzi Shan, Ngozi Osuji, Jie Ding, Nino Isakadze, Francoise A Marvel, Garima Sharma, Seth S Martin. Originally published in JMIR Cardiol (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 Cardiol, 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.