Validation of an algorithm for continuous monitoring of atrial fibrillation using a consumer smartwatch

Robert Avram, MD, MSc,* Mattheus Ramsis, MD,* Ashley D. Cristal, MD,* Viswam Nathan, PhD,† Li Zhu, PhD,† Jacob Kim, MS,† Jilong Kuang, PhD,† Alex Gao, MS,† Eric Vittinghoff, PhD,‡ Linnea Rohdin-Bibby, BS,* Sara Yogi, BA,* Emina Seremet, BS,* Valerie Carp, BS,* Fabio Badilini, PhD,§ Mark J. Pletcher, MD, MPH,‡ Gregory M. Marcus, MD, MAS, FHRS,* David Mortara, PhD,§ Jeffrey E. Olgin, MD, FHRS*

*From the *Division of Cardiology, University of California San Francisco, San Francisco, California, ‡Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, §Center for Physiologic Research, School of Nursing, University of California San Francisco, San Francisco, California, and †Samsung Research America, Mountain View, California.*

BACKGROUND Consumer devices with broad reach may be useful in screening for atrial fibrillation (AF) in appropriate populations. However, currently no consumer devices are capable of continuous monitoring for AF.

OBJECTIVE The purpose of this study was to estimate the sensitivity and specificity of a smartwatch algorithm for continuous detection of AF from sinus rhythm in a free-living setting.

METHODS We studied a commercially available smartwatch with photoplethysmography (W-PPG) and electrocardiogram (W-ECG) capabilities. We validated a novel W-PPG algorithm combined with a W-ECG algorithm in a free-living setting, and compared the results to those of a 28-day continuous ECG patch (P-ECG).

RESULTS A total of 204 participants completed the free-living study, recording 81,944 hours with both P-ECG and smartwatch measurements. We found sensitivity of 87.8% (95% confidence interval [CI] 83.6%–91.0%) and specificity of 97.4% (95% CI 97.1%–97.7%) for the W-PPG algorithm (every 5-minute classification); sensitivity of 98.9% (95% CI 98.1%–99.4%) and specificity of 99.3% (95% CI 99.1%–99.5%) for the W-ECG algorithm; and sensitivity of 96.9% (95% CI 93.7%–98.5%) and specificity of 99.3% (95% CI 98.4%–99.7%) for W-PPG triggered W-ECG with a single W-ECG required for confirmation of AF. We found a very strong correlation of W-PPG in quantifying AF burden compared to P-ECG ($r = 0.98$).

CONCLUSION Our findings demonstrate that a novel algorithm using a commercially available smartwatch can continuously detect AF with excellent performance and that confirmation with W-ECG further enhances specificity. In addition, our W-PPG algorithm can estimate AF burden. Further research is needed to determine whether this algorithm is useful in screening for AF in select at-risk patients.

KEYWORDS Atrial fibrillation; Atrial fibrillation burden; Photoplethysmography Remote monitoring; Screening; Smartwatch

(Heart Rhythm 2021;18:1482–1490)

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, affecting more than 33 million adults worldwide.[1] Approximately one-third of individuals who have AF are asymptomatic.[2] The most serious complication of AF is cardioembolic stroke, with one-third of all strokes attributed to AF. Moreover, the amount of AF (defined as AF burden) and not just the presence of AF is related to risk of stroke and development of heart failure.[3,4] Therefore, an approach to large-scale screening of AF in asymptomatic participants at risk for stroke, as well as monitoring of AF burden, could have a major impact. Furthermore, having an approach that is tunable to the risk and prevalence of AF is crucial to minimize user burden and unnecessary utilization of health care resources.

Baseline screening of AF in high-risk individuals can lead to fewer patient-years with undetected AF, fewer strokes, and

Funding sources: This research was funded by a grant from Samsung to Dr Olgin (Principal Investigator). Drs Olgin, Pletcher, and Marcus are funded by National Institutes of Health Grant U2CEB021881. Disclosures: Dr Nathan, Dr Zhu, Jacob Kim, Dr Kuang, and Alex Gao are employees of Samsung Research America. All other author have reported that they have no relationships relevant to the contents of this paper to disclose. The entire study was performed by and all data (including from the smartwatch) were stored, managed, and analyzed by the UCSF team. **Address reprint requests and correspondence:** Dr Jeffrey E. Olgin, University of California San Francisco, Division of Cardiology, 505 Parnassus Ave, Box 0124, San Francisco, CA 94939. E-mail address: Jeffrey.olgin@ucsf.edu.

1547-5271/

https://doi.org/10.1016/j.hrthm.2021.03.044

overall increase in life-years in participants at high risk for stroke.[5,6] Mobile health technology, specifically noninvasive wearables with photoplethysmography (PPG), has been a promising avenue for the detection of AF and may be an effective and practical approach for large-scale screening given their growing wide-spread use.[7–12]

Although consumer-grade on-demand electrocardiogram (ECG) devices have been developed and used to detect symptomatic AF, they have minimal utility for detecting asymptomatic AF. We first reported a machine learning algorithm to detect AF from heart rate measurements derived from a commercial smartwatch,[10] and a similar approach has been used to trigger an on-demand ECG.[13] The Apple Heart Study demonstrated the utility of a simple noncontinuous tachogram measure of PPG irregularity that monitors for 1 minute every 2 hours (and only when the user is still).[14] Importantly, none of these approaches are robust enough to allow for true near-continuous monitoring to quantify the amount of AF (ie, AF burden).

The objective of our study was to develop and validate a robust algorithm to detect AF episodes from sinus rhythm (SR) and quantify AF burden in a real-world setting. This was done using the Samsung Galaxy Watch Active 2 (Samsung, Seoul, South Korea), a consumer-grade device equipped with a continuous watch photoplethysmography (W-PPG) sensor and an on-demand watch ECG (W-ECG) sensor. We report the performance of this algorithm in a remote, real-world setting and compare it to the gold-standard continuous 28-day ECG patch monitor.

Methods

The research study was conducted with approval of the University of California San Francisco (UCSF) Institutional Review Board. Informed consent was obtained from all participants.

W-PPG and W-ECG algorithm development

Details of the W-PPG and W-ECG algorithm development are described in the Supplemental Appendix and Supplemental Table 1.

Free-living validation study

After the W-PPG and W-ECG algorithms had been developed and optimized, we set out to validate them in a real-world setting. We enrolled participants with a known prior diagnosis of AF or those at risk for developing AF in a remote-based study (My Heart Lab Validation Study; ClinicalTrials.gov Identifier: NCT04314947) in which participants simultaneously wore a continuous ECG patch (Biotel ePatch, BioTelemetry, Inc., Malvern, PA) and a Samsung watch for 4 consecutive weeks.

Study population

To achieve the targeted enrollment of 200 participants contributing data for analysis, participants were screened and enrolled from April 2020 to June 2020 from the Health eHeart Study, a large online observational cohort of >300,000. The study was conducted completely remotely using the Eureka Research Platform (https://info.eurekaplatform.org/), a research platform developed in part through National Institutes of Health funding for conducting mobile research, supporting over 40 mobile studies with >400,000 participants. Participants were screened for a self-reported diagnosis of AF or with any one of the following risk factors: age 65 years and older, or history of hypertension, diabetes, heart failure, or coronary disease. Eligible participants were invited to participate in the MyHeartLab Validation study via an e-mail containing a link to the study in the Eureka Research mobile application (app).

Enrollment and data collection

All study-related activities were performed on the participants' own smartphone using the Eureka Research app. After patients confirmed their eligibility and provided their consent and shipping address within the app, a study kit consisting of two 14-day BioTel ePatches (P-ECG) and a smartwatch were shipped to the participant. Participants completed surveys to report their demographics, relevant medical history, AF symptom burden, and medications.[15] To increase adherence to the study activities, regular reminders using push notifications and SMS text reminders were sent via the Eureka Platform. Two weeks after completion of the first patch, the participant received a notification and instructions to change the ePatch. At the end of the study, participants received automated reminders to remove and mail back the ePatches, and the watch was remotely wiped of the investigational software.

The Samsung Galaxy Active 2 smartwatch used in this study is a consumer-grade device with the hardware to obtain continuous W-PPG recordings for AF detection and an on-demand, single-lead W-ECG. The watch required a specific investigational software app that was preloaded and used only in this study to activate the hardware and run the algorithms. Using the algorithm described in the Supplemental Appendix, the W-PPG continuously monitored heart rhythm and rate, and classified a nonoverlapping, moving 5-minute window into "regular," "irregular," or "uncertain." If the W-PPG detected an irregular rhythm over a 5-minute window, a notification was sent to the participant to take an on-demand W-ECG (Figure 1). If the W-ECG was not done, repeat notifications were sent after 15 minutes, 1 hour, 2 hours, and 4 hours. If the first W-ECG showed AF, the participant was reminded in 1 hour to take another W-ECG. If the W-ECG results in an "inconclusive" classification, participants were reminded to repeat the W-ECG in 5 minutes. Participants were also randomly notified to record 1 W-ECG per day and were able to record a W-ECG on demand.

Data management and processing

Upon return of the ePatches, the data were uploaded for processing, analysis, and adjudication. All P-ECG signals were converted to ISHNE format for input into the UCSF's Signal

Processing Core using CER-S (Continuous ECG Recording Suite; AMPS-LLC, NY, NY) for further processing and analysis. AF episodes were identified and overread by a technician and a board-certified cardiologist. All periods of SR (defined as heart rate <100 bpm with P waves prior to every QRS; <4 premature atrial contractions or premature ventricular contractions in a row; and no bigeminy or trigeminy) were identified. Episodes of AF were reviewed to ensure that they did not overlap with SR episodes. Participants with ≥1% ventricular pacing were excluded from the analysis since the algorithm was not designed to detect AF during regular ventricular pacing.

All signals and data from the Samsung smartwatch was collected via WiFi or cellular service and stored in the Eureka Platform through an Application Program Interface. The W-ECG and W-PPG signals were time-aligned with the P-ECG. Because we were interested in evaluating the performance of the algorithm in distinguishing AF from SR, we analyzed all SR and AF P-ECGs that had simultaneous W-PPG or W-ECG recordings. Furthermore, 3000 randomly sampled W-ECGs (ensuring at least 1/3 sampled are classified as SR) from the pool of recorded W-ECGs were adjudicated by 2 board-certified cardiologists (with a third electrophysiologist breaking any ties). Finally, we calculated AF burden (percent time in AF) using the W-PPG algorithm and compared it to that determined by P-ECG.

Statistical analysis

Descriptive statistics are reported as count (percentage) for categorical variables and median (interquartile range [IQR]) or mean ± SD for continuous variables, as labeled. Sensitivity, specificity, and area under the curve for the detection of AF episodes were calculated using the AF or SR result on the P-ECG as the gold standard. To account for repeated measures and estimate confidence intervals (CIs) of sensitivity and specificity, we used intercept-only generalized linear mixed logistic models for test positivity or negativity, restricted to true-positive or true-negative intervals, respectively. To assess the accuracy of AF burden measurement based on W-PPG in reference to P-ECG, a Pearson correlation with mean squared error was performed.

Results

E-mail invitations were sent to 10,042 Health eHeart Study participants, of whom 40% (4003) opened the e-mail and 7.7% (783) clicked on the link to learn about the study. A total of 332 participants completed the eligibility survey, of whom 16 were not eligible for the study (Figure 2). Of the 295 consented participants, the first 221 participants received the study kit, and the other 74 participants were put on a waiting list. Fourteen participants did not contribute data: 12 due to a technical problem in the first batch of watches that prevented data transmission and 2 due to unreturned patches. Therefore, a total 207 participants contributed data from both the ePatch and smartwatch (Figure 2). After manual review, 2 participants with significant (>90%) ventricular pacing and 1 who had poor P-ECG signal quality were excluded from analysis. Thus, we report data from 204 participants who had 384 ePatches analyzed and a total of 5462 participant-days of data.

Characteristics of the participants are listed in Table 1. Participants wore the smartwatch a median of 20.9 hours/day (IQR 18.8–21.9). The W-PPG monitored participants for an average of 19.5 ± 4.2 hours/day for a total of 106,663 hours. A total of 32.2% of the W-PPG data was classified as indeterminate, 7.4% as AF, and 60.2% as SR. Participants on average took 5.4 ± 3.7 W-ECGs per day, totaling 24,209 W-ECGs recorded during the study period (Supplemental Figure 1). The median rate of recording a W-ECG in response to a W-PPG alert was 66.7% (IQR 34.4%–100%). The indeterminant rate for W-ECG rhythms recorded during periods of SR on the P-ECG was 52.3% (1.0% due to rate cutoffs, 2.2% due to ectopic beats, and 49.1% due to poor signal quality) and for those that were recorded during AF on P-ECG was 43.7% (0.1% due to rate cutoffs, 0.7% due to ectopic beats, and 42.9% due to poor signal quality).

A total of 81,944 hours of monitoring from the ePatch with simultaneous W-PPG data was recorded and analyzed. There were 266 AF episodes ≥5 minutes in 54 participants and 200 episodes ≥60 minute in 53 participants recorded on the P-ECG while simultaneously wearing the smartwatch. The time of day of W-PPG classifications and W-ECGs taken are shown in Figure 3. There were 145 AF episodes detected on a single W-ECG and 95 episodes confirmed on a second W-ECG taken at least 1 hour after the first W-ECG, with 6 episodes without a W-ECG recorded at all.

The performances of the W-PPG and W-ECG algorithms are listed in Table 2. Sensitivity and specificity of the continuous 5-minute (nonoverlapping moving window) W-PPG

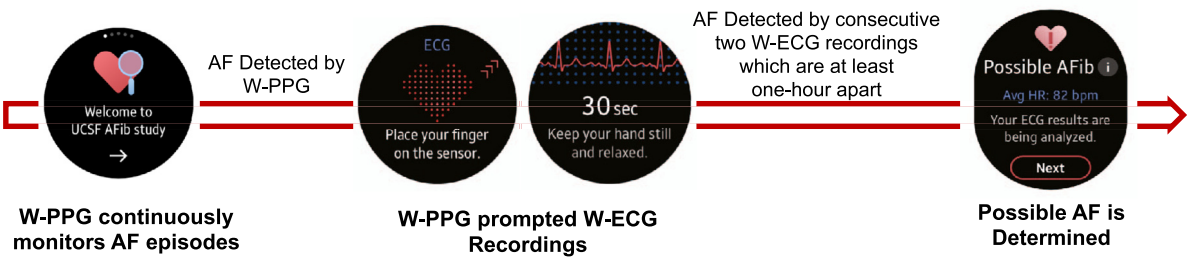

Figure 1 Screenshots and flow of notifications on the smartwatch in the investigational software. AF = atrial fibrillation; AFib = atrial fibrillation; ECG = electrocardiogram; HR = heart rate; W-ECG = watch electrocardiography; W-PPG = watch photoplethysmography.

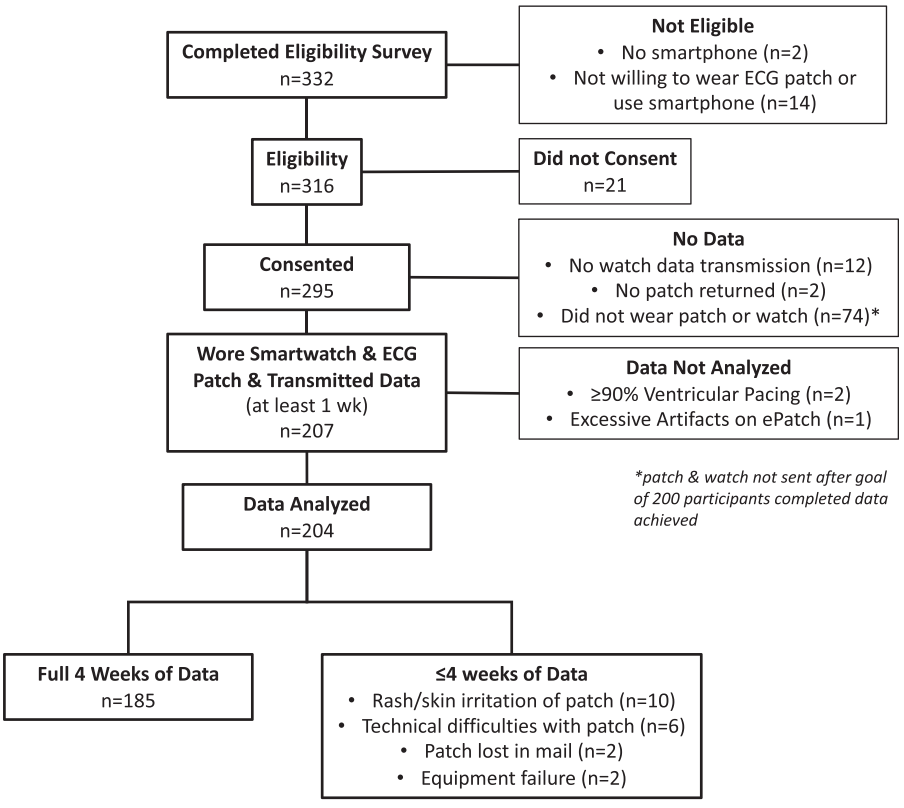

Figure 2 Consort diagram showing flow of screening, enrollment, and data analyses in the study. ECG = electrocardiogram.

decisions alone were 87.8% (95% CI 83.6%–91.0%) and 97.4% (95% CI 97.1%–97.7%), respectively. We also determined the performance of an *a priori* scheme for notifying users to check a W-ECG based on 2-of-3, 5-minute W-PPG classifications meeting AF criteria. This *a priori* notification-level performance demonstrated sensitivity of 81.8% (95% CI 71.2%–89.1%) and specificity of 99.4% (95% CI 99.3%–99.5%). The performances of the algorithm when other rhythms were included (in addition to AF and SR) found on P-ECG are listed in Supplemental Table 2 and are nearly identical to those when only AF and SR on P-ECG were analyzed. Figure 4 shows the performance at the notification level across the spectrum of n–1/n 5-minute decisions to trigger a user notification.

Sensitivity and specificity of the W-ECG algorithm alone compared to the P-ECG were 98.9% (95% CI 98.1%–99.4%) and 99.3% (95% CI 99.1%–99.5%), respectively. Using the random sample of W-ECGs adjudicated by 2 cardiologists, sensitivity of the W-ECG was 98.8% and specificity 97.0% compared to expert rhythm interpretation. When assessing the performance of W-PPG triggered W-ECG with a single W-ECG required for confirmation, sensitivity and specificity were 96.9% (95% CI 93.7%–98.5%) and 99.3% (95% CI 98.4%–99.7%), respectively. When a second W-PPG triggered W-ECG done at least 1 hour apart from the first W-ECG was required, sensitivity decreased slightly to 96.5% (95% CI 94.1%–97.9 %) but specificity increased to 100% (95% CI: 100%–100%).

W-PPG performance in quantifying AF burden compared to P-ECG was excellent (Figure 5), with correlation of $r = 0.98$ and mean average error of 4.3% ± 4.6%. The error in estimating AF burden was highest for longer AF burden.

Discussion

We observed a high accuracy of a novel continuous W-PPG algorithm for monitoring of irregular heart rhythms combined with W-ECG detection of AF in a free-living setting using a smartwatch compared to a 28-day ePatch (P-ECG). In addition, the watch's performance in measuring AF burden by continuous W-PPG was excellent. To our knowledge, this is the first study to compare the accuracy of a W-PPG triggered W-ECG in a free-living setting to detect AF, which can reduce the number of false-positive diagnoses, and the first to measure AF burden using a smartwatch.

We demonstrated that our smartwatch-derived PPG has excellent performance characteristics, with high sensitivity (87.8%) and specificity (97.4%). Previous studies demonstrated that PPG algorithms had sensitivity between 90% and 96% and specificity between 85% and 99% for diagnosis of AF.[7,8,16–21] Despite this excellent performance, these studies followed participants for a limited amount of time and usually only looked at short recordings of PPG of <1 hour and in a controlled setting. In MyHeartLab, we collected >106,663 hours of simultaneous watch and P-ECG data and report on the performance of every

Table 1 Characteristics of the study population (N = 204)

Age (yr)	62.61 ± 11.60
<65 yr	101 (52)
65–75 yr	72 (37)
>75 yr	23 (11)
Sex	
Male	112 (54)
Female	95 (46)
Race	
Caucasian	185 (89)
Hispanic	10 (5)
African-American	8 (4)
Asian	8 (4)
History of AF	
No history	32 (15)
Paroxysmal	159 (77)
Persistent	16 (8)
Symptoms of AF	
In AF at time of study	29 (15)
Daily	10 (5)
Weekly	50 (26)
Monthly	40 (20)
Within 1 yr	40 (20)
>1 yr	14 (7)
Never aware	13 (7)
Comorbidities	
History of HTN	106 (51)
History of CHF	33 (16)
Previous MI	20 (10)
History of coronary disease	42 (20)
Prior CVA or TIA	20 (10)
Diabetes	30 (14)
Obstructive sleep apnea	85 (41)
CHA_2DS_2VaSc score	
0–1	64 (32)
2–4	115 (58)
5–7	17 (8)

Values are given as mean ± SD or n (%).

AF = atrial fibrillation; CHF = congestive heart failure; CVA = cerebrovascular accident; HTN = hypertension; MI = myocardial infarction; TIA = transient ischemic attack.

classification of the algorithm. Tison et al[10] looked at time series of heart rate values (not using the raw PPG signal) collected using the Apple Watch for the detection of AF to develop and validate a deep neural network algorithm. The algorithm demonstrated excellent performance against 12-lead ECG (sensitivity 98.0%; specificity 90.2%) for the detection of AF when tested in a controlled environment. However, when deployed in a free-living setting, performance decreased (sensitivity 67.7%; specificity 67.6%). This approach was also limited by the lack of "access" to the raw PPG signals as well as signals from other sensors, thus limiting the algorithm to use of only averaged heart rate data. Our W-PPG algorithm—based on a heuristic signal processing approach using the raw PPG signal and inputs from other sensors—performed better than the machine learning approach based on heart rate. The Apple Heart Study also evaluated the performance of a smartwatch PPG to detect irregular rhythms.[14] The study reported a positive predictive value of the PPG algorithm of 71% but did not report the sensitivity or specificity of the algorithm or the negative predictive value of the algorithm. For comparison, the W-PPG algorithm reported herein had a positive predictive value of 82.3% for a single 5-minute classification and 88.7% for the 2-of-3, 5-minute classification notification scheme. Importantly, the algorithm reported in the Apple Heart Study was not a continuous monitor (measuring only 1 minute every 2 hours when a user is completely still) nor was it combined with an on-demand W-ECG.[14]

In our study, we used the W-PPG as a passive monitor for irregular heart rhythms, which can be used to then trigger a notification prompting the participant to record a confirmatory W-ECG. Using this approach, we had high sensitivity (96.9%) and specificity (99.3%). The specificity can be even further improved (100%) by requiring a second confirmatory W-ECG taken at least 1 hour apart while having only minimal impact on sensitivity (96.0%). No other published study has evaluated the performance of combining a robust continuous PPG monitor from a smartwatch triggering a confirmatory ECG. This may make the approach attractive in screening the general population in which very high specificity is desired and detection of AF episodes >1 hour is most appropriate.

The algorithm developed and tested in this study also has the advantage that it can easily be tuned to optimize specificity for populations in which the prevalence of AF is low to avoid false-positive results or sensitivity (still with a good specificity) in populations in which the prevalence is higher and the risk of missing the diagnosis is higher (eg, elderly patients with cryptogenic stroke or those with symptoms consistent with AF). By increasing the number of required 5-minute AF "decisions" to just 5-of-6 by the W-PPG algorithm required to detect AF, specificity of 99.8% can be achieved. In contrast, in high-prevalence populations, sensitivity can be increased by reducing the number of 5-minute decisions required to detect AF. Our data suggest that PPG alone may be sufficient for monitoring AF, especially in the evening when the rate of undetermined rhythms are quite low or in populations with prior diagnosis or high prevalence.

Optimizing specificity, either by increasing the number of required "AF decisions" before notifying the patient or by requiring confirmatory W-ECGs, is critical to minimize burden and utilization of health care system resources for false-positive results. Current wearables suffer from high false-positive notification rates, and, importantly, true test performance for these devices has not been reported in the literature. For example, although the Apple Watch has approval from the US Food and Drug Administration for their W-PPG and W-ECG technology, the device has not been adequately studied, and postmarket surveillance demonstrated a high false-positive rate, with only 15% of patients who received an abnormal pulse alert on an Apple Watch having AF confirmed.[22]

Importantly, by incorporating data from the raw PPG signals and the other sensors in the watch, we were able to

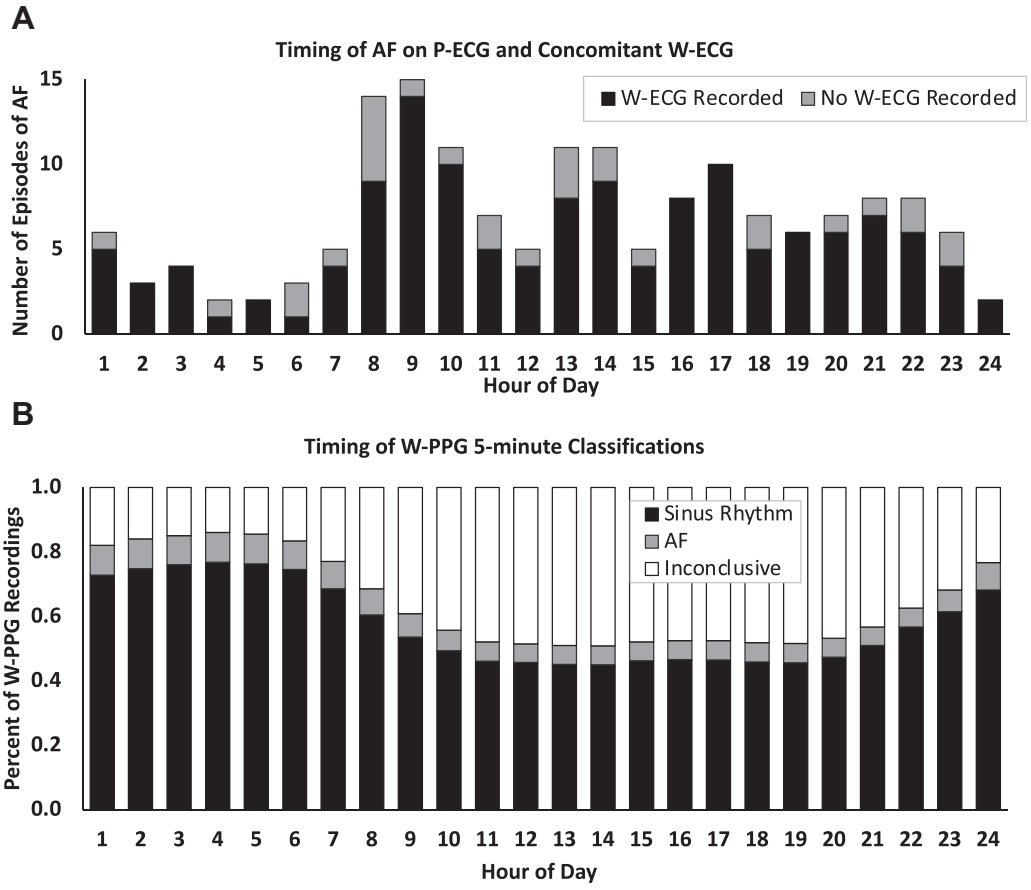

Figure 3 **A:** Plot of the number of AF episodes detected on patch electrocardiography (P-ECG) by time of day and W-ECG classification. **B:** Stacked bar plot of the distribution of W-PPG classifications according to the hour of the day. Abbreviations as in Figure 1.

Table 2 Algorithm performance

	SENS	$SENS_{est}$ (95% CI)	SPEC	$SPEC_{est}$ (95% CI)	AUC	TP (N)	TN (N)
W-PPG performance							
Algorithm level (every 5-min classification)	89.7	87.8 (83.6–91.0)	97.0	97.4 (97.1–97.7)	93.3	54,754	590,438
Notification level (2-of-3 classifications)	90.8	81.8 (71.2–89.1)	98.8	99.4 (99.3–99.5)	94.8	55,431	601,577
W-ECG performance							
Algorithm level compared to P-ECG	97.4	98.9 (98.1–99.4)	96.5	99.3 (99.1–99.5)	97.1	1,712	10,259
Algorithm level compared to rhythm adjudication by cardiologists	98.8	–	97.0	–	96.0	584	835
W-PPG triggered W-ECG performance							
Diagnosis confirmed by a single W-ECG	96.9	96.9 (93.7–98.5)	99.3	99.7 (99.5–99.7)	97.7	219	51,456
Diagnosis confirmed by 2 W-ECGs ≥1 h apart*	96.0	96.0 (92.9–97.8)	100	100 (100–100)	98.0	265	74,248
Daily participant performance							
W-PPG triggered W-ECG confirmation (2 W-ECGs ≥1 h apart) detection of AF or SR per day*	96.9	96.9 (93.9–98.4)	99.4	99.9 (99.7–100)	98.2	251	3772

Values are given as % unless otherwise indicated.

AF = atrial fibrillation; AUC = area under the receiver operator characteristic curve; CI = confidence interval; P-ECG = patch electrocardiography; SENS = sensitivity; $SENS_{est}$ = estimated sensitivity using logistic regression; SPEC = specificity; $SPEC_{est}$ = estimated sensitivity using logistic regression; SR = sinus rhythm; TN = number of true-negative events (SR on P-ECG); TP = number of true-positive events (AF on P-ECG); W-ECG watch electrocardiography; W-PPG = watch photoplethysmography.

*Includes only episodes lasting ≥60 min on P-ECG. Also includes those having confirmation with 2 subsequent W-ECGs classified as AF (or one as SR) even if the first W-ECG was classified as undetermined.

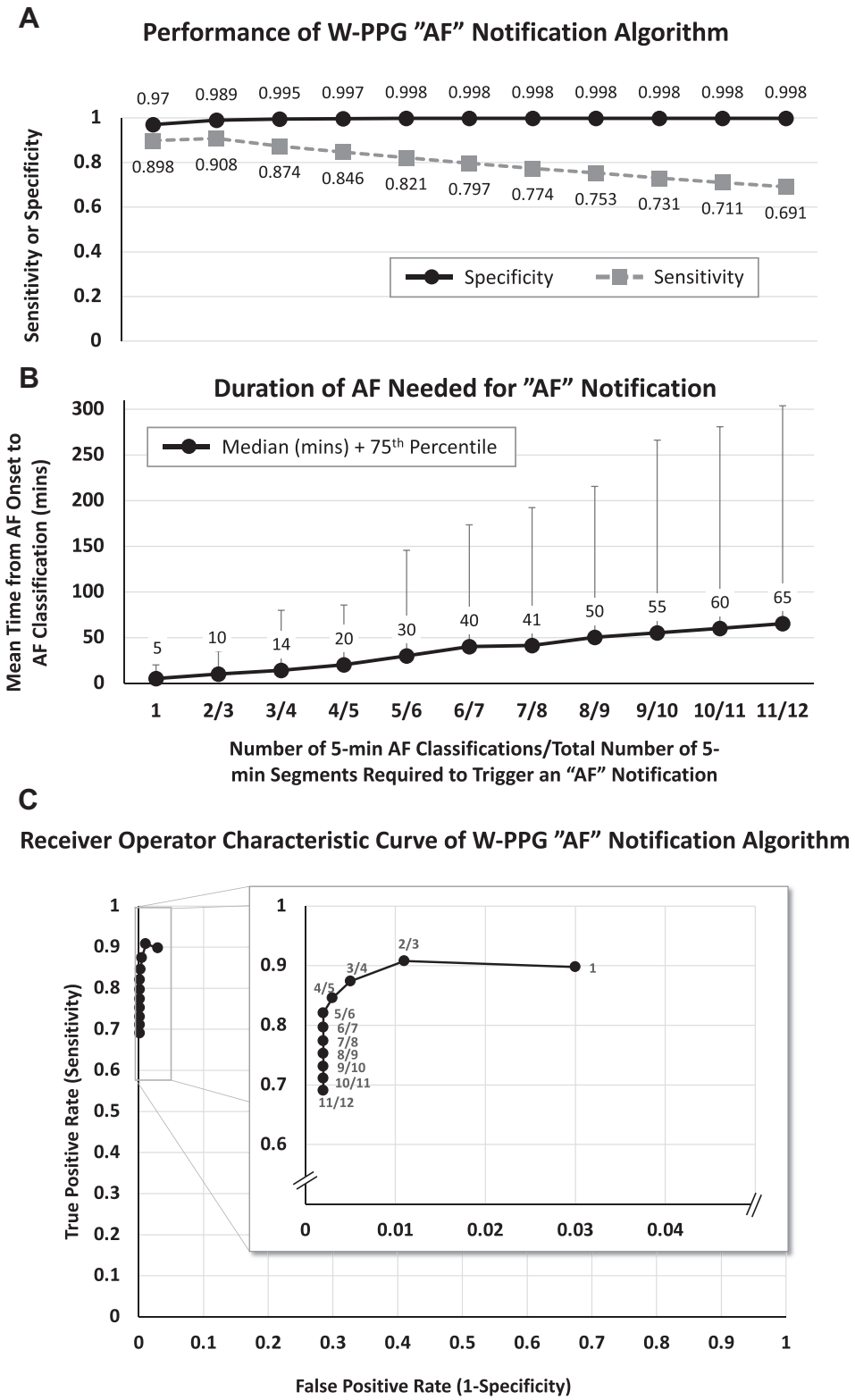

Figure 4 **A:** Specificity and sensitivity of the W-PPG AF notification algorithm (notification of the user of possible AF) with increasing the required number of 5-minute AF classifications over an increasing period of time (n–1/n). Data include all episodes of AF longer than (n–1) × 5 minutes at each point. **B:** Median time after onset AF (from P-ECG) from first 5 minutes of AF to meeting criteria for notifying the user of possible AF with increasing the required number of 5-minute AF classifications over an increasing period of time (n–1/n). Error bar represents the 75th percentile. **C:** Receiver operator characteristic (ROC) curve across the range of increasing 5-minute AF classifications (n–1/n) required for notifying for AF. *Inset* represents zoomed in view of the ROC curve with truncated axes. Abbreviations as in Figure 1.

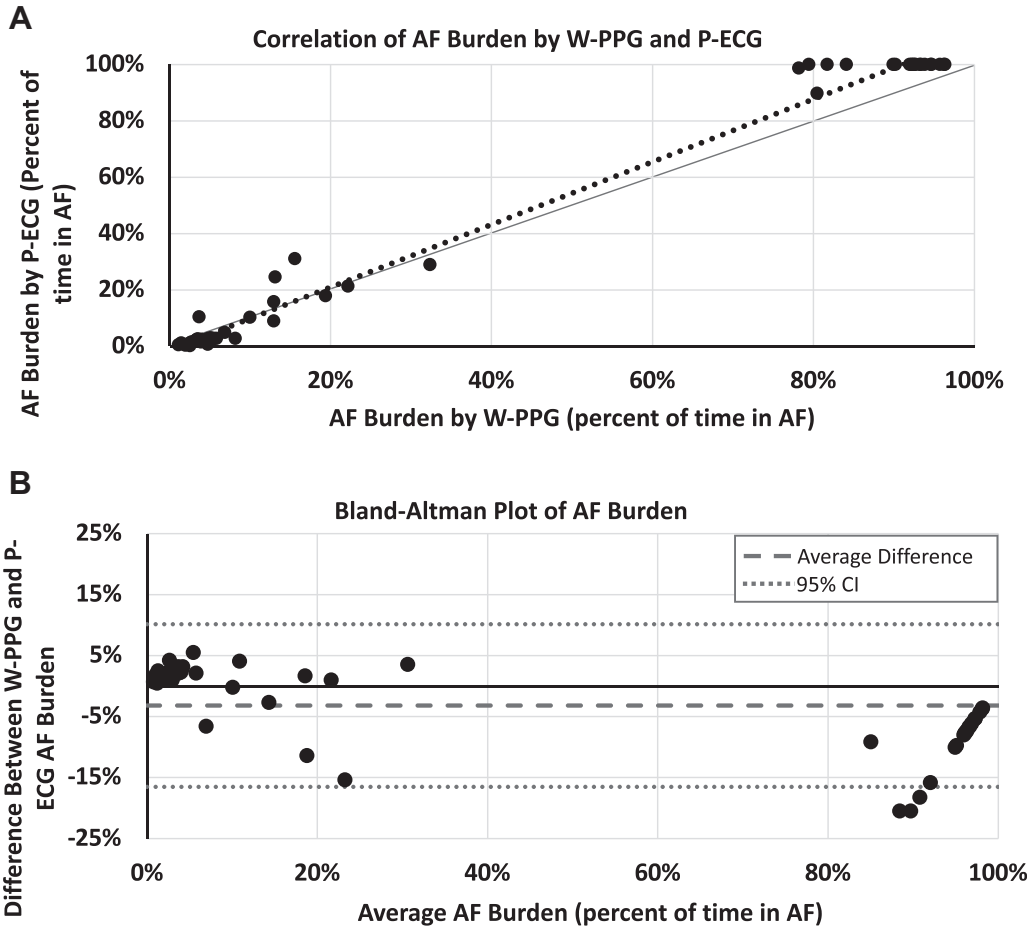

Figure 5 **A:** Scatter plot and regression line of AF burden (%) estimated by W-PPG compared to AF burden (%) estimated by P-ECG, showing a correlation of r = 0.98. Each *dot* represents the AF burden by each method for each patient with AF. *Dashed line* is the regression line. *Solid gray line* is line of equality. **B:** Bland-Altman analysis showed good agreement between quantification of AF burden on W-PPG and P-ECG. *Dots* represent the difference in AF burden (%) between that determined by W-PPG and that determined by P-ECG for each participant with AF. *Dashed line* is the average value of difference between AF burden (%) by W-PPG and P-ECG. *Dotted line* is the upper and lower 95% confidence interval (CI) of the difference between AF burden (%) by W-PPG and P-ECG. Abbreviations as in Figures 1 and 3

develop a W-PPG algorithm that provided monitoring over most of the day. Over the median 20.9 hours/day the smartwatch was worn, an average of 6.7 hours (non-continuous) of recorded data was classified as inconclusive by the W-PPG algorithm. This compares to other consumer devices in which monitoring for an irregular rhythm occurs for 1 minute every 2 hours,[14] amounting to a total of 4 minutes of monitoring during a typical day of wearing the watch. As a result of this near-continuous monitoring, we were able to monitor AF burden using continuous W-PPG monitoring for the first time in a consumer-grade watch with a high correlation to that on P-ECG. Studies have demonstrated that AF burden is directly related to an increased risk of cardioembolic stroke and heart failure.[5,13,23] Although most clinicians utilize ECG patches for this purpose, their utility is limited because they are short-term recordings and cannot be easily deployed in real time. A smartwatch has a multiday battery life (and can be fully recharged in a few hours) and could allow for measurement of AF burden easily over a very long period of time.

Study limitations

The majority of the MyHeartLab participants were Caucasian. Although we did not see differences in signals across races and previous data suggest that these devices perform similarly across the spectrum of Fitzpatrick skin tones,[24] current findings may be insufficient to apply to people with different skin tones. Importantly, the algorithms were not designed to detect rhythms other than AF or SR. Notably, our population had a high burden of other rhythms as well as ectopy and still performed reasonably well for the detection of AF and SR. As shown in Supplemental Table 2, when we include rhythms other than AF and SR (as determined on P-ECG) in the analyses, the performance is nearly identical to that listed in Table 2. We did not assess the impact of an algorithm on AF screening or utilization of health care resources; such studies will need to be completed to assess the safety and efficacy of such an algorithm for screening and management of AF. The W-ECG algorithm limited rhythm decisions to rates >50 bpm and <110 bpm to avoid misdiagnosing other more serious arrhythmias and thus may

be less useful in patients with AF and rapid ventricular rates. This also may have contributed to a lower sensitivity. The algorithm does not adequately detect or quantify AF in patients with concurrent ventricular pacing or atrial flutter, as the algorithm detects the pulse as either regular or irregular, regardless of the true underlying electrophysiological rhythm.

Conclusion

Continuous monitoring W-PPG, on-demand W-ECG sensor, and W-PPG triggered W-ECG recording demonstrated excellent diagnostic accuracy for AF. Our findings suggest that a robust smartwatch algorithm can limit the number of false-positive alerts when used in a free-living setting. Furthermore, our algorithm on the smartwatch was able to determine AF burden reliably. More studies are needed to assess its performance as a screening tool for AF and its impact on utilization of health care resources.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hrthm.2021.03.044.

References

1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation 2014;129:837–847.
2. Rienstra M, Lubitz SA, Mahida S, et al. Symptoms and functional status of patients with atrial fibrillation: state of the art and future research opportunities. Circulation 2012;125:2933–2943.
3. Chen LY, Chung MK, Allen LA, et al. Atrial fibrillation burden: moving beyond atrial fibrillation as a binary entity: a scientific statement from the American Heart Association. Circulation 2018;137:e623–e644.
4. Lin HJ, Wolf PA, Benjamin EJ, Belanger AJ, D'Agostino RB. Newly diagnosed atrial fibrillation and acute stroke. The Framingham Study. Stroke 1995;26:1527–1530.
5. Anter E, Jessup M, Callans DJ. Atrial fibrillation and heart failure: treatment considerations for a dual epidemic. Circulation 2009;119:2516–2525.
6. Levin L, Husberg M, Sobocinski PD, et al. A cost-effectiveness analysis of screening for silent atrial fibrillation after ischaemic stroke. Europace 2015;17:207–214.
7. Nemati S, Ghassemi MM, Ambai V, et al. Monitoring and detecting atrial fibrillation using wearable technology. Annu Int Conf IEEE Eng Med Biol Soc 2016;2016:3394–3397.
8. Bumgarner JM, Lambert CT, Hussein AA, et al. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018;71:2381–2388.
9. Raja JM, Elsakr C, Roman S, et al. Apple Watch, wearables, and heart rhythm: where do we stand? Ann Transl Med 2019;7:417.
10. Tison GH, Sanchez JM, Ballinger B, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018;3:409–416.
11. Inui T, Kohno H, Kawasaki Y, et al. Use of a smart watch for early detection of paroxysmal atrial fibrillation: validation study. JMIR Cardio 2020;4:e14857.
12. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. Circulation 2015;131:2176–2184.
13. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019;12:e006834.
14. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909–1917.
15. Spertus J, Dorian P, Bubien R, et al. Development and validation of the Atrial Fibrillation Effect on QualiTy-of-Life (AFEQT) questionnaire in patients with atrial fibrillation. Circ Arrhythm Electrophysiol 2011;4:15–25.
16. Dörr M, Nohturfft V, Brasier N, et al. The WATCH AF Trial: SmartWATCHes for Detection of Atrial Fibrillation. JACC Clin Electrophysiol 2019;5:199–208.
17. Koenig N, Seeck A, Eckstein J, et al. Validation of a new heart rate measurement algorithm for fingertip recording of video signals with smartphones. Telemed J E Health 2016;22:631–636.
18. Krivoshei L, Weber S, Burkard T, et al. Smart detection of atrial fibrillation†. Europace 2017;19:753–757.
19. Chan PH, Wong CK, Poh YC, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. J Am Heart Assoc 2016;5:e003428.
20. McManus DD, Lee J, Maitas O, et al. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. Heart Rhythm 2013;10:315–319.
21. Brasier N, Raichle CJ, Dörr M, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Europace 2019;21:41–47.
22. Wyatt KD, Poole LR, Mullan AF, Kopecky SL, Heaton HA. Clinical evaluation and diagnostic yield following evaluation of abnormal pulse detected using Apple Watch. J Am Med Inform Assoc 2020;27:1359–1363.
23. Van Gelder IC, Healey JS, Crijns H, et al. Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. Eur Heart J 2017;38:1339–1344.
24. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18.