

# Continuous atrial fibrillation monitoring using a wearable smartwatch: Using long-term Holter as reference

DIGITAL HEALTH  
Volume 11: 1–10  
© The Author(s) 2025  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076251314105  
journals.sagepub.com/home/dhj



Yannan Pan<sup>1,\*</sup> , Erdong Chen<sup>1,\*</sup> , Shihui Jie<sup>1</sup>, Dongbo Huo<sup>1</sup>,  
Zhongru Ding<sup>2</sup>, Jing Zhou<sup>1</sup>, Jie Jiang<sup>1</sup> , Jianping Li<sup>1</sup> and Yong Huo<sup>1</sup>

## Abstract

**Background:** Wearables satisfactorily detect atrial fibrillation (AF) longer than 1 hour. Our study aims to evaluate smartwatch performances for long-term AF monitoring, including AF with short durations.

**Methods:** This prospective study enrolled AF patients from 2020 to 2023. Diagnostic efficacy of the Amazfit smartwatch, with AF-identifying algorithms from photoplethysmography (PPG) and single-lead electrocardiogram (ECG), was compared with a 7-day Holter. Primary analysis included smartwatch diagnostics to identify AF longer than 5 minutes. Secondary analyses evaluated smartwatch performances under different settings and compared AF burdens between the smartwatch and Holter.

**Results:** The study analyzed 72 patients (48 males, mean age  $65.4 \pm 8.5$ ) with 914 AF episodes lasting 834.7 hours, including 142 longer-than-5-minute AF episodes. Smartwatch recording time was 8927.6 hours. By individual, sensitivities and specificities of AF longer than 5 minutes were 100.0% and 83.7% for PPG and 89.7% and 67.4% for the ECG algorithm. Positive and negative predictive values were 94.9% and 99.9% for PPG and 77.6% and 99.8% for ECG. Optimal AF durations to be identified by PPG and ECG algorithms were 1.358 and 16.708 minutes. Smartwatch performances varied across AF durations and between day-time and night-time. Strong correlations (PPG:  $\rho = 0.877$ ; ECG:  $\rho = 0.769$ ) and excellent agreements (PPG: ICC = 0.976; ECG: ICC = 0.927) were found between AF burdens calculated from smartwatch and Holter.

**Conclusions:** Compared with long-term Holter, the wearable smartwatch had satisfying qualitative and quantitative diagnostic performances for continuous AF monitoring. Susceptibility to false positives led to modest specificity. Smartwatch performances were affected by AF durations and time periods.

**Registration:** ChiCTR2000040035.

## Keywords

Cardiovascular disease, cardiology, artificial intelligence, digital health, wearables

Submission date: 6 August 2024; Acceptance date: 3 January 2025

## Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia worldwide and is associated with increased risks of stroke, heart failure, cognitive decline, hospitalization, and death.<sup>1</sup> Screening for AF is challenging due to its occult, paroxysmal, and asymptomatic characteristics. Traditional 12-lead electrocardiograms (ECG) or Holter ECG are unavailable for long-term continuous

<sup>1</sup>Department of Cardiology, Peking University First Hospital, Beijing, China  
<sup>2</sup>Huami (Beijing) Information Technology Co. Ltd, Beijing, China

\*These authors contributed equally to this work.

### Corresponding author:

Jie Jiang, Department of Cardiology, Peking University First Hospital, No.8 Xishiku Street, Xicheng District, Beijing, 100034, China.  
Email: jiangjie417@vip.163.com



monitoring, while implantable loop recorders are less adaptive due to invasiveness and high costs.<sup>2</sup> Wearable devices are therefore desirable for their convenience and accuracy,<sup>3</sup> with increasing uses of commercially available devices.

The Amazfit Health Smartwatch is a wearable device with a photoplethysmography (PPG) sensor and a single-lead ECG (SL-ECG) recorder,<sup>4</sup> with algorithms to classify rhythm strips. A previous study demonstrated that the Amazfit wearable had a satisfactory performance for single time point AF detection, with an accuracy of 93.27% and 94.76% for PPG and ECG readings.<sup>4</sup>

Few studies evaluated smartwatch performances for continuous AF monitoring. Two studies used an insertable cardiac monitor (ICM) or cardiac implanted electronic device (CIED) as references and found the smartwatch could satisfactorily detect AF longer than 1 hour.<sup>5,6</sup> However, endpoints were AF episodes lasting longer than an hour because ICM was susceptible to false positives for shorter AF duration.<sup>7</sup> A few studies used conventional Holter as the reference to evaluate smartwatch performances,<sup>8,9</sup> and more comprehensive results are still warranted. Our study also chose long-term Holter as the gold standard to identify AF more accurately and comprehensively, including those with short durations. The study aimed to: (1) investigate the diagnostic efficacy of a wearable smartwatch for long-term continuous AF monitoring based on PPG and ECG signals in a real-world environment, using long-term Holter ECG as references; (2) evaluate smartwatch performances under different settings; (3) estimate AF burdens using a wearable smartwatch.

## Methods

### Study design

This prospective, single-center study aimed to evaluate the diagnostic performances of a wearable smartwatch. Patients were recruited from Peking University First Hospital, a tertiary hospital in China, between November 2020 and March 2023. The study enrolled participants across inpatient and outpatient settings, with a minimum age requirement of 18. The study population was restricted to AF patients who met at least one of the inclusion criteria: (1) patients admitted to inpatient wards for AF catheter ablation and were in sinus rhythm at discharge; (2) paroxysmal or persistent AF patients from outpatient settings who were in sinus rhythm at study enrollment. Patients were excluded if they had implanted pacemakers, were allergic to wearable smartwatches, or were unable to use smartwatches in bilateral upper extremities due to any of the following conditions: bilateral upper limb disabilities, abnormal skin lesions on wrists, obliterative vascular diseases of upper extremities, or severe upper extremity edema.

Written informed consent was obtained from all study individuals, and the study was approved by the institutional

ethics committee of Peking University First Hospital (2020-312). The study was registered in the Chinese Clinical Trial Registry (Registration number: ChiCTR2000040035). The investigation was performed according to the principles outlined in the Declaration of Helsinki. The research reported in this paper adhered to the Standards for Reporting Diagnostic Accuracy (STARD) statement.<sup>10</sup>

### Study device

The Amazfit Health Smartwatch is a wearable device that integrates signals from PPG sensors or single-channel ECG recorders with AF-identifying algorithms using a deep convolutional neural network.<sup>4</sup> PPG signals are continuously and automatically collected for 71 seconds for each PPG segment, while SL-ECG signals are manually recorded for 60 seconds for each ECG segment. Each PPG or ECG segment is categorized into the following algorithmic outputs: AF, not AF, poor signal quality, and unable to identify. Samples are shown in Supplemental Figure S1. In this study, a long-term Holter ECG monitor was used as the reference. Details of the smartwatch and Holter monitor are shown in Supplemental Methods S1.

### Study procedure

Baseline characteristics were collected, and all individuals underwent standardized 12-lead ECG at enrollment to verify heart rhythm. Patients wore the smartwatch for 7 days and completed a 7-day Holter monitor test simultaneously. Time systems were standardized by comparing SL-ECG recordings from the smartwatch and Holter. During the study period, patients were encouraged to record smartwatch SL-ECG 4 times per day (8:00, 12:00, 16:00, 20:00) and any time they felt palpitations or discomforts. SL-ECG recording compliance of each patient was categorized into regularly recorded (recorded more than 85% as instructed) and intermittently recorded (recorded less than 85% as instructed).

Two experienced ECG physicians manually reviewed long-term Holter recordings independently, blinded to patient characteristics and smartwatch recordings. An AF episode was defined as irregular RR intervals without discernible repeated P waves lasting longer than 30 seconds. Any disagreements in AF identification were verified by an electrophysiologist. For each AF episode, precise starting and ending times were recorded, and AF duration was calculated. Time periods of poor signal quality or no recognizable recordings from Holter were excluded from the study analysis.

For each PPG or ECG segment, recording time and its algorithmic output were extracted from the web-based platform. Individuals who did not wear the smartwatch as instructed with a smartwatch recording time less than

20% of the 7-day study period, were excluded from the analysis.

### Data analysis

In this study, we chose 5 minutes as the cutoff duration threshold of clinically meaningful AF, following recommendations from most studies that examined the relationship between AF burdens and outcomes.<sup>11</sup> The primary analysis focused on smartwatch performance to detect clinically meaningful AF from an individual, daily, or episode-based approach. Diagnostic sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were calculated, with detailed definitions in Supplemental Methods S2. For individual-based and daily-based analyses, diagnostic efficacies were evaluated using PPG, ECG, and two combinations of PPG and ECG algorithms. In these combinations, positive results were defined as AF detection in either PPG or ECG or in both PPG and ECG.

Secondary analyses included: (1) diagnostic efficacy of the smartwatch from a segment-based analysis, with definitions demonstrated in Supplemental Methods S2; (2) causes of the false positive and false negative segments; (3) smartwatch performances for different AF durations; (4) PPG diagnostics according to daily time periods, including day-time from 7:00 to 23:00 and night-time from 23:00 to 7:00; (5) comparison of AF burdens between the smartwatch and Holter. AF burden from Holter was defined as the total duration of AF divided by the total analyzable time of Holter, while AF burden from the smartwatch was defined as the number of AF segments divided by the sum of AF and non-AF segments.

### Statistical analysis

Categorical variables were presented with counts and percentages. Continuous variables were reported as means and standard deviations, or medians and interquartile ranges (IQR) if skewed distribution, while distributions were confirmed by the Shapiro–Wilk normality test. For diagnostic sensitivity, specificity, accuracy, NPV, and PPV, results were presented as percentages and 95% confidence intervals (CI). Agreements between the smartwatch and Holter were compared using Cohen’s kappa coefficient, with greater than 0.75 representing excellent agreement, 0.40–0.75 indicating fair to good agreement, and less than 0.40 representing poor agreement.<sup>12</sup> Diagnostic efficacies across different settings were compared using chi-square or Fisher’s exact tests, as appropriate. For AF burdens, skewed distributions were confirmed, and Spearman’s rank correlation coefficient and intraclass correlation coefficient (ICC) were used to assess correlation and agreement between AF burdens calculated from the smartwatch and

**Table 1.** Smartwatch performances for longer than 5-minute AF detection.

	Holter			
	Individual-based analysis		Daily-based analysis	
	Positive	Negative	Positive	Negative
PPG				
Positive	29	7	92	27
Negative	0	36	7	422
ECG				
Positive	26	14	69	36
Negative	3	29	30	413
PPG or ECG <sup>a</sup>				
Positive	29	19	97	56
Negative	0	24	2	393
PPG and ECG <sup>b</sup>				
Positive	26	2	64	7
Negative	3	41	35	442

Smartwatch performances are evaluated from individual-based and daily-based analyses, using Holter ECG as references.

<sup>a</sup>Positive results are defined as AF detection in either PPG or ECG.

<sup>b</sup>Positive results are defined as AF detection in both PPG and ECG.

AF: atrial fibrillation, ECG: electrocardiogram, PPG: photoplethysmography.

Holter. Receiver operating characteristic (ROC) curve analysis was used to find the optimal AF duration and AF burden thresholds to be identified by the smartwatch. The Bonferroni correction was used for multiple comparisons, and a *P*-value less than .05 divided by the number of comparisons was considered significant. Statistical analyses were performed with R software version 4.1.2.

Sample size estimation was based on a one-sided exact test to detect superiority to a target accuracy of 70%, with an assumed smartwatch accuracy of 85%, considering previous single time point diagnostic performance of the Amazfit wearable and the impact of artifacts during continuous smartwatch monitoring.<sup>4,13</sup> Sixty-one patients were needed to achieve a statistical power of 80% and a one-sided significance level of 0.025. The sample size was calculated using PASS Version 15.0.5 (NCSS, LLC, Kaysville, Utah, USA).

**Table 2.** Smartwatch diagnostic efficacy for AF lasting longer than 5 minutes.

	Sensitivity	Specificity	Accuracy	PPV	NPV	Cohen's kappa coefficient
Individual-based analysis						
PPG	100.0 (88.1–100.0)	83.7 (69.3–93.2)	90.3 (81.0–96.0)	80.6 (64.0–91.8)	100.0 (90.3–100.0)	0.806
ECG	89.7 (72.6–97.8)	67.4 (51.5–80.9)	76.4 (64.9–85.6)	65.0 (48.3–79.4)	90.6 (75.0–98.0)	0.538
PPG or ECG <sup>a</sup>	100.0 (88.1–100.0)	55.8 (39.9–70.9)	73.6 (61.9–83.3)	60.4 (45.3–74.2)	100.0 (85.8–100.0)	0.504
PPG and ECG <sup>b</sup>	89.7 (72.6–97.8)	95.3 (84.2–99.4)	93.1 (84.5–97.7)	92.9 (76.5–99.1)	93.2 (81.3–98.6)	0.855
Daily-based analysis						
PPG	92.9 (86.0–97.1)	94.0 (91.4–96.0)	93.8 (91.4–95.7)	77.3 (68.7–84.5)	98.4 (96.7–99.3)	0.806
ECG	69.7 (59.6–78.5)	92.0 (89.1–94.3)	88.0 (84.9–90.6)	65.7 (55.8–74.7)	93.2 (90.5–95.4)	0.603
PPG or ECG <sup>a</sup>	98.0 (92.9–99.8)	87.5 (84.1–90.4)	89.4 (86.5–91.9)	63.4 (55.2–71.0)	99.5 (98.2–99.9)	0.705
PPG and ECG <sup>b</sup>	64.6 (54.4–74.0)	98.4 (96.8–99.4)	92.3 (89.8–94.4)	90.1 (80.7–95.9)	92.7 (89.9–94.8)	0.709

Results are presented with % (95% CI). Smartwatch diagnostic efficacies are evaluated from individual-based and daily-based analyses.

<sup>a</sup>Positive results are defined as AF detection in either PPG or ECG.

<sup>b</sup>Positive results are defined as AF detection in both PPG and ECG.

AF: atrial fibrillation, CI: confidence interval, ECG: electrocardiogram, NPV: negative predictive value, PPG: photoplethysmography, PPV: positive predictive value.

## Results

### Study population and baseline characteristics

The study enrolled 78 eligible patients, and 72 patients were included in the final analysis, with 6 excluded due to a smartwatch recording time of less than 20% of the study period. A detailed enrollment flowchart is shown in Supplemental Figure S2. Baseline characteristics are shown in Supplemental Table S1. The mean age was 65.4 ± 8.5 years, with 48 (66.7%) male patients. The study enrolled 18 (25.0%) inpatient post-ablation AF patients and 54 (75.0%) AF patients from outpatient settings.

The median wear time for Holter was 168.1 (IQR: 168.0–168.3) hours per individual. After subtracting those times with no or poor-quality signals, the median analyzable time for Holter was 157.0 (IQR: 134.1–166.1) hours per individual. Thirty-four (47.2%) patients had documented AF from Holter, with 29 (40.3%) having AF lasting longer than 5 minutes. AF duration was 834.7 hours in total, with an average AF burden of 7.9%. The total number of AF episodes was 914, including 142 lasting longer than 5 minutes. The median AF duration was 1.0 (IQR: 0.7–2.3) minutes for each AF episode and 23.8 (IQR: 9.2–107.1) minutes for each clinically meaningful AF episode. For smartwatch statistics, the total smartwatch recording time was 8927.6 hours, with a median of 130.8

(IQR: 115.1–146.1) hours per individual and 18.7 (IQR: 16.4–20.9) hours per individual per day. The median numbers of PPG and ECG segments per individual were 6630 (IQR: 5836–7407) and 34.5 (IQR: 27–47). For SL-ECG recording compliance, 62 patients (86.1%) regularly recorded SL-ECG as instructed, while 10 patients (13.9%) intermittently followed the study procedure. No adverse events were reported during the study process.

### Primary analysis of smartwatch diagnostic efficacy

Smartwatch performances and diagnostic efficacy for AF longer than 5 minutes are shown in Tables 1 and 2. From an individual-based approach, sensitivities and specificities for AF detection were 100.0% (95% CI, 88.1–100.0%) and 83.7% (95% CI, 69.3–93.2%) for PPG and 89.7% (95% CI, 72.6–97.8%) and 67.4% (95% CI, 51.5–80.9%) for ECG algorithms. Combining PPG and ECG results, sensitivity could be improved to 100.0% (95% CI, 88.1–100.0%) when positive results were defined as AF detection in either PPG or ECG, while specificity could be improved to 95.3% (95% CI, 84.2–99.4%) when positive results were defined as AF detection in both PPG and ECG.

From a daily-based approach, sensitivities and specificities for AF detection were 92.9% (95% CI, 86.0–97.1%) and 94.0% (95% CI, 91.4–96.0%) for PPG and 69.7%

(95% CI, 59.6–78.5%) and 92.0% (95% CI, 89.1–94.3%) for ECG algorithms.

From an episode-based analysis, 99 (69.7%) episodes were identified as AF from PPG, and 43 (30.3%) were missed, with 4 (2.8%) due to smartwatch non-wearing. For ECG algorithms, 48 (33.8%) episodes were identified as AF, and 94 (66.2%) were missed, with 92 (64.8%) due to non-documentation. Sensitivities of PPG and ECG algorithms were 69.7% (95% CI, 61.5–77.1%) and 33.8% (95% CI, 26.1–42.2%) if non-wearing were regarded as negative

results, while 71.7% (95% CI, 63.5–79.1%) and 96.0% (95% CI, 86.3–99.5%) if non-wearing were excluded.

### Segment-based analysis and causes of false readings

Smartwatch performances by segment-based analyses are shown in Table 3. The numbers and percentages of AF, not AF, poor signal quality, and unable to identify segments were 16,473 (3.6%), 230,437 (50.9%), 201,673 (44.6%), 4083 (0.9%) for PPG algorithms, and 331 (12.4%), 1996 (74.7%), 316 (11.8%), 28 (1.0%) for ECG algorithms. Diagnostic efficacy from segment-based analysis is shown in Table 4.

Of the 835 false positive PPG segments, 570 (68.3%) were due to atrial flutter with variable conduction, 136 (16.3%) were due to frequent premature atrial contractions or atrial tachycardia, and 129 (15.4%) were due to sinus arrhythmia. Among the 73 false positive ECG segments, 45 (61.6%) were due to signal artifacts, 18 (24.7%) were due to atrial flutter with variable conduction, 5 (6.8%) were due to sinus arrhythmia, 3 (4.1%) were due to frequent premature atrial contractions and atrial tachycardia, and the other 2 (2.7%) were in sinus rhythm without identifiable reasons. Samples of false positives are shown in Supplemental Figure S3. Both of the 207 false negative PPG segments and 3 false negative ECG segments were in AF rhythm.

### Smartwatch performances according to AF durations

Sensitivities of smartwatch AF detection across different AF durations are shown in Figure 1. Sensitivities for AF episodes lasting from 0.5 to 5 minutes, 5 to 60 minutes, 1 to 24 hours, and longer than 24 hours were 23.3% (95% CI, 20.1–26.7%), 64.8% (95% CI, 53.9–74.7%), 81.8% (95% CI, 67.3–91.8%), and 100.0% (95% CI, 54.1–100.0%) by PPG, while 0.5% (95% CI, 0.1–1.3%), 10.2% (95% CI, 4.8–18.5%), 72.7% (95% CI, 57.2–85.0%), and 100.0% (95% CI, 54.1–100.0%) by smartwatch ECG.

**Table 3.** Smartwatch performances by PPG and ECG segment-based analyses.

	Holter		Transitions between AF and non-AF
	AF	Not AF	
PPG algorithm			
AF	15,434	835	204
Not AF	207	229,846	384
Poor signal quality	18,525	182,763	385
Unable to identify	2853	962	268
ECG algorithm			
AF	253	73	5
Not AF	3	1990	3
Poor signal quality	19	295	2
Unable to identify	2	26	0

AF: atrial fibrillation, ECG: electrocardiogram, PPG: photoplethysmography.

**Table 4.** Smartwatch diagnostic efficacy by PPG and ECG segment-based analyses.

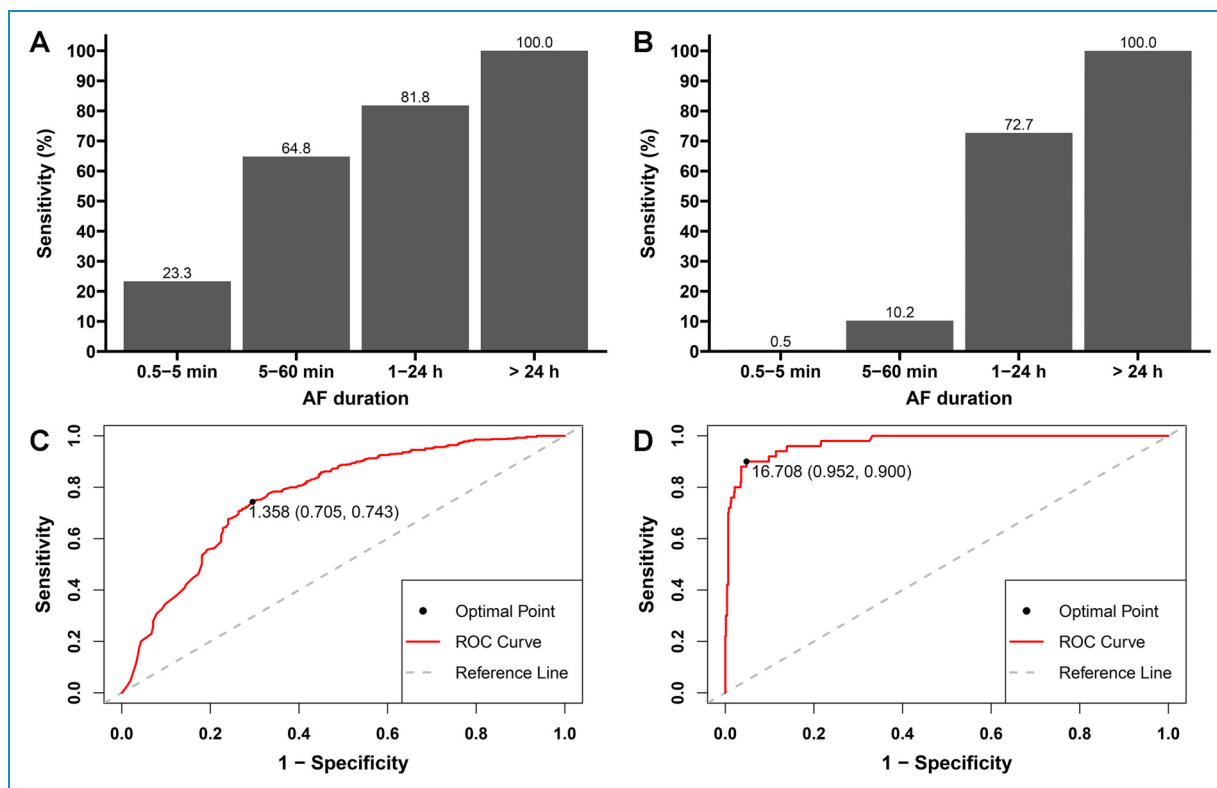
	Sensitivity, including unrecognizable segments <sup>a</sup>	Sensitivity, excluding unrecognizable segments <sup>b</sup>	PPV	Specificity, including unrecognizable segments <sup>a</sup>	Specificity, excluding unrecognizable segments <sup>b</sup>	NPV
PPG	41.7 (41.2–42.2)	98.7 (98.5–98.8)	94.9 (94.5–95.2)	55.5 (55.3–55.6)	99.6 (99.6–99.7)	99.9 (99.9–99.9)
ECG	91.3 (87.4–94.4)	98.8 (96.6–99.8)	77.6 (72.7–82.0)	83.5 (81.9–84.9)	96.5 (95.6–97.2)	99.8 (99.6–100.0)

Results are presented with % (95% CI). For AF detection, sensitivity and PPV are calculated. For non-AF detection, specificity and NPV are calculated.

<sup>a</sup>Poor signal quality and unable to identify segments were included and regarded as negative results.

<sup>b</sup>Poor signal quality and unable to identify segments were excluded from the analysis.

AF: atrial fibrillation, CI: confidence interval, ECG: electrocardiogram, NPV: negative predictive value, PPG: photoplethysmography, PPV: positive predictive value.



**Figure 1.** Smartwatch performances according to AF durations. Sensitivities of PPG (A) and smartwatch ECG (B) for AF lasting from 0.5 to 5 minutes, 5 to 60 minutes, 1 to 24 hours, and longer than 24 hours. ROC curve analyses demonstrate optimal AF duration thresholds to be identified by PPG (C) and smartwatch ECG (D) are 1.358 and 16.708 minutes. AF: atrial fibrillation, ECG: electrocardiogram, PPG: photoplethysmography, ROC: receiver operating characteristic.

Significant differences were found in sensitivities across different AF durations ( $P < .001$ ). ROC curve analyses are shown in Figure 1, demonstrating that 1.358 and 16.708 minutes were optimal AF duration thresholds to be identified by PPG and ECG algorithms.

### PPG diagnostics by daily time periods

Segment-based PPG diagnostics according to daily time periods are demonstrated in Figure 2 and compared between day-time and night-time in Supplemental Table S2. Significant differences ( $P < .001$ ) were found in sensitivity, specificity, and accuracy, which were likely caused by different proportions of poor signal quality between day-time and night-time. PPV and NPV were satisfying all day long.

### Smartwatch AF burden calculation

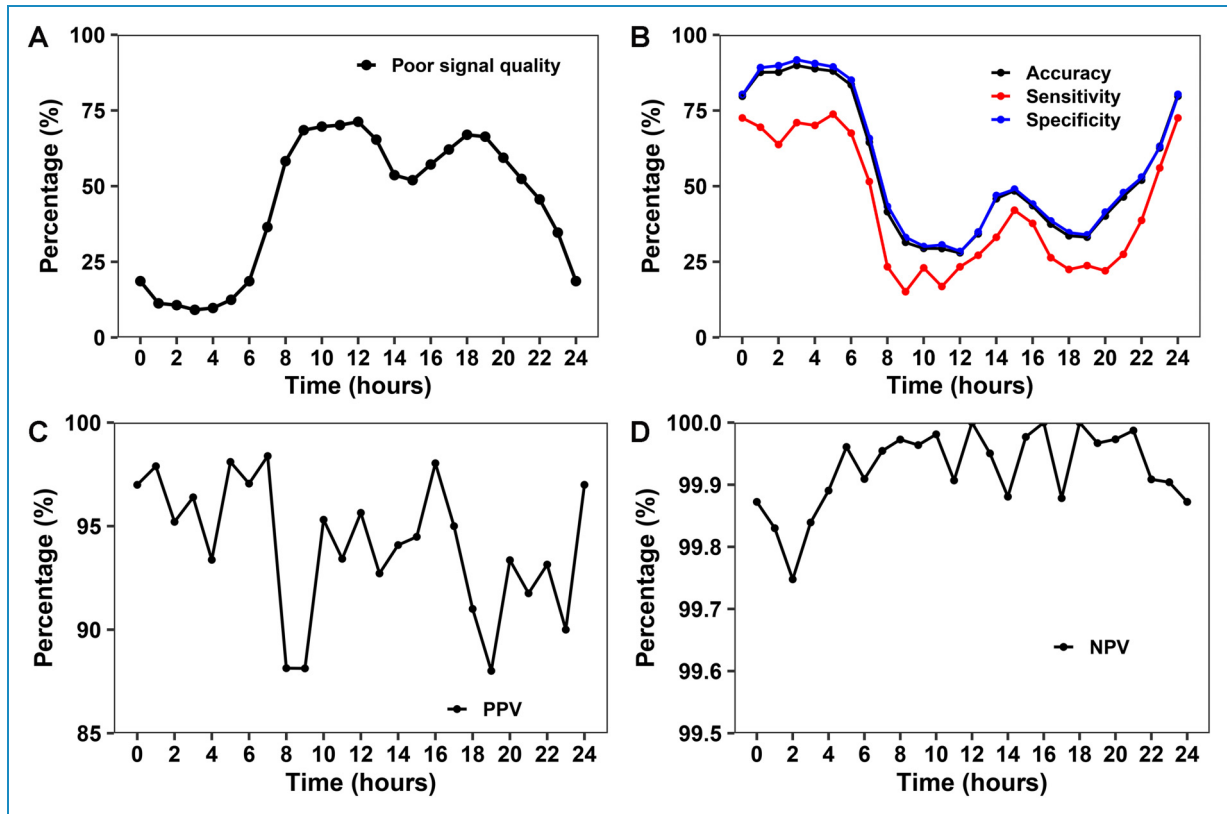
Correlations between AF burdens calculated from smartwatch PPG or ECG and Holter are shown in Figure 3. A stronger correlation was found in PPG ( $\rho = 0.877$ ,  $P < .001$ ) than in smartwatch ECG ( $\rho = 0.769$ ,  $P < .001$ ). Agreements between AF burdens calculated from the

smartwatch and Holter were excellent, with ICCs of 0.976 (95% CI, 0.962–0.985) and 0.927 (95% CI, 0.815–0.964) for PPG and ECG algorithms. ROC curve analyses demonstrated that 0.401% and 1.339% were optimal AF burden thresholds to be identified by PPG and ECG algorithms, as shown in Supplemental Figure S4.

### Discussion

In this prospective study of AF patients, the Amazfit Health Smartwatch demonstrated well-performed sensitivities but modest specificities for AF detection. The PPG algorithms outperformed smartwatch SL-ECG, and combinations of algorithms could separately improve sensitivity or specificity. From an episode-based analysis, the primary cause of detection failure was smartwatch non-wearing or non-recording. Both algorithms had excellent PPV for AF and NPV for non-AF.

A preliminary study on single time point diagnostic performances showed the Amazfit wearable's promising ability to detect AF.<sup>4</sup> Statistics of smartwatch performances in a real-world ambulatory setting are still warranted. Two studies used ICM or CIED as references to evaluate smartwatch performances to detect AF longer than 1 hour.<sup>5,6</sup> Our



**Figure 2.** PPG diagnostics according to daily time periods. Variations of PPG performances according to daily time periods, including the percentage of poor signal quality segments (A), accuracy, sensitivity, specificity (B), PPV (C), and NPV (D). NPV: negative predictive value, PPG: photoplethysmography, PPV: positive predictive value.

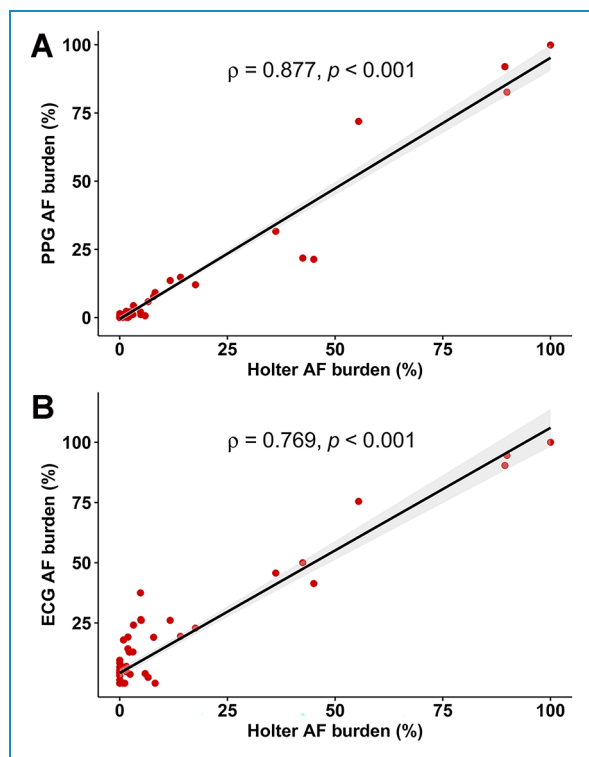
study used long-term Holter as references and shortened the target AF duration to 5 minutes. Our research examined the diagnostic efficacy from multiple perspectives and through different settings. For AF burdens, despite SL-ECG being collected intermittently and PPG being susceptible to unanalyzable segments, strong correlations and agreements were still found between AF burdens calculated from the smartwatch and Holter, corroborating previous findings<sup>14,15</sup> and supporting future use of smartwatch to estimate AF burdens in real-world settings.

PPG has the advantage of passive and continuous monitoring to identify shorter AF episodes with better diagnostic efficacy, compared with intermittently recorded smartwatch ECG. Despite the merits, PPG is vulnerable to motion artifacts and has the drawback of high proportions of unanalyzable data due to poor signal quality.<sup>13</sup> In our study, nearly half of the PPG outputs were inconclusive due to poor signal quality or inability to identify. Sensitivity and specificity were excellent when excluding these inconclusive segments, while substantially minimized if included. Therefore, improving PPG signal quality is crucial for enhancing smartwatch performance. Similar to a previous study,<sup>14</sup> variations in proportions of unanalyzable segments between day-time and night-time were found. This could be explained by the

associations between PPG analyzability and day-time activity level.<sup>16</sup> For AF screening or detection, the smartwatch may have better diagnostic efficacy for nocturnal AF.

In our study, both algorithms were susceptible to false-positive outputs, which led to the modest specificity results during individual-based analysis. PPG had lower false-positive rates than smartwatch ECG, and the ECG algorithm had greater percentages of false positives since ECG segment analysis highly depends on the quality of ECG tracings. False positives during AF screening could cause anxiety and unnecessary costs, and the SMART WARS study recommended clinician interpretations of SL-ECG to improve accuracy.<sup>17</sup> In our research, combining PPG and ECG results could minimize the percentage of false positives and increase specificity.

For AF patients, AF burden correlates with clinical outcomes, including risks of stroke, heart failure, and mortality.<sup>18</sup> Since many AF patients are asymptomatic, it is difficult to estimate AF burden without monitors. The smartwatch could facilitate AF burden measurements and indicate prognostic values. Moreover, wearables could be used for long-term rhythm monitoring in post-ablation AF patients.<sup>19</sup> For the general population, the smartwatch could be used to screen for subclinical AF. Plenty of



**Figure 3.** Scatterplots of AF burdens calculated from the smartwatch and Holter. Scatterplots of AF burdens calculated from Holter and PPG (A) or smartwatch ECG (B). Regression lines are in black with gray areas representing 95% confidence intervals. Spearman's rank correlation coefficients are calculated. AF: atrial fibrillation, ECG: electrocardiogram, PPG: photoplethysmography.

studies have focused on the associations between subclinical AF durations and outcomes. However, no consensus has been reached on the optimal cutoff of subclinical AF duration that impacts prognosis, ranging from 1 minute to 24 hours.<sup>11</sup> Longer durations were correlated with increased risks of adverse events, and most studies found that subclinical AF lasting longer than 24 hours was associated with the highest risks of adverse outcomes, including stroke and major adverse cardiovascular events.<sup>20–23</sup> In our study, the smartwatch performed better in identifying longer AF episodes. Therefore, the smartwatch could identify those subclinical AF episodes with worse prognostic values, while the missed episodes have little impact on prognosis. It is noteworthy that these studies were based on subclinical AF found by CIEDs, and future results of episodes detected by wearables are needed to confirm the relationship between subclinical AF duration and prognosis.

Despite these promising clinical implications, physicians have considerable variations in responses to AF alerts detected by smartwatches,<sup>24</sup> and no randomized trials have proven the prognostic benefits of AF screening from commercialized devices.<sup>25</sup> Standardized recommendations

for smartwatch-detected AF should be promoted. Future randomized studies are needed to determine whether the prospective use of smartwatches to monitor AF burden or screen subclinical AF could improve the prognosis. Moreover, how to enhance smartwatch performances with shorter AF durations and during movements brings a challenge to future artificial intelligence algorithms.

### Limitations

The study provided valuable results on smartwatch ECG performances. However, smartwatch ECG performances highly rely on measurement frequency and the symptomatic nature of AF, which could lead to potential bias. Moreover, PPV and NPV results were affected by the percentage of the population who had AF during the study. Since more than 40% of our study population had AF episodes, the PPV and NPV results should be taken cautiously when generalizing smartwatch performances to the general population, with a much lower prevalence rate of AF. Smartwatch non-documentation or unanalyzable segments affect diagnostic performances. The average AF burden in this study was 7.9%, high enough for some of the AF episodes to be recognized despite non-documentation or unanalyzable segments. When generalizing the study results to populations with lower AF burdens, unanalyzable segments or smartwatch non-recording might have a greater impact on diagnostic performances. The most prominent risk factor for AF is advancing age,<sup>26</sup> and elderly patients are less adaptive to digital health. It should be noted that the study population was relatively young, with a mean age of 65.4. More data is needed on whether the results apply to the real-world AF population with a greater percentage of elderly patients. Lastly, most patients wore the smartwatch for more than 18 hours daily. In real-world settings, patient noncompliance could lead to less wear time and decreased diagnostic performance. Future studies examining the diagnostic efficacy through longer study periods are needed.

### Conclusions

Using long-term Holter as the reference, wearable smartwatch had satisfying qualitative and quantitative diagnostic performances for continuous AF detection, while susceptibility to false positives led to modest specificity. PPG was susceptible to unanalyzable data, and its diagnostic performances varied according to measuring time periods. The smartwatch was quantitatively accurate for AF monitoring and could be used to estimate AF burdens in clinical AF patients.

**Contributorship:** Study design: EC and JJ. Software and data analysis: YP. ECG interpretation: EC, SJ, and YP. Original draft: YP. Editing and revision: EC, SJ, DH, ZD, JZ, JJ, JL, and



YH. Study supervision: JJ. All authors have read and approved the final version of the manuscript.

**Data availability:** All the data associated with this study are available from the corresponding author upon reasonable request.


**Declaration of conflicting interests:** The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Zhongru Ding is the Medical Director of Huami (Beijing) Information Technology Co. Ltd. All the other authors have no conflicts of interest to disclose.

**Ethical approval and informed consent:** The study was approved by the institutional ethics committee of Peking University First Hospital (2020-312). Written informed consent was obtained from all study individuals. The investigation was performed according to the principles outlined in the Declaration of Helsinki.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was sponsored by a grant from Huami (Beijing) Information Technology Co. Ltd.

**Guarantor:** JJ.

**ORCID iDs:** Yannan Pan  <https://orcid.org/0000-0003-4868-6649>

Erdong Chen  <https://orcid.org/0000-0002-3401-1512>

Jie Jiang  <https://orcid.org/0000-0002-4077-0056>

**Supplemental material:** Supplemental material for this article is available online.

## References

- Hindricks G, Potpara T, Dagres N, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European association for cardio-thoracic surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European heart rhythm association (EHRA) of the ESC. *Eur Heart J* 2021; 42: 373–498. 2020/08/30.
- Giancaterino S, Lupercio F, Nishimura M, et al. Current and future use of insertable cardiac monitors. *J Am Coll Cardiol EP* 2018; 4: 1383–1396. 2018/11/24.
- Hermans ANL, Gawalko M, Dohmen L, et al. Mobile health solutions for atrial fibrillation detection and management: a systematic review. *Clin Res Cardiol* 2022; 111: 479–491. 2021/09/23.
- Chen E, Jiang J, Su R, et al. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. *Heart Rhythm* 2020; 17: 847–853. 2020/05/02.
- Wasserlauf J, Vogel K, Whisler C, et al. Accuracy of the Apple watch for detection of AF: A multicenter experience. *J Cardiovasc Electrophysiol* 2023; 34: 1103–1107. 2023/03/22.
- Wasserlauf J, You C, Patel R, et al. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol* 2019; 12: e006834. 2019/05/23.
- Mittal S, Rogers J, Sarkar S, et al. Real-world performance of an enhanced atrial fibrillation detection algorithm in an insertable cardiac monitor. *Heart Rhythm* 2016; 13: 1624–1630. 2016/05/12.
- Reissenberger P, Serfözö P, Piper D, et al. Determine atrial fibrillation burden with a photoplethysmographic mobile sensor: the atrial fibrillation burden trial: detection and quantification of episodes of atrial fibrillation using a cloud analytics service connected to a wearable with photoplethysmographic sensor. *Eur Heart J Digit Health* 2023; 4: 402–410. 2023/10/05.
- Chang PC, Wen MS, Chou CC, et al. Atrial fibrillation detection using ambulatory smartwatch photoplethysmography and validation with simultaneous holter recording. *Am Heart J* 2022; 247: 55–62. 2022/02/09.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Br Med J* 2015; 351: h5527. 2015/10/30.
- Kreimer F, Mügge A and Gotzmann M. How should I treat patients with subclinical atrial fibrillation and atrial high-rate episodes? Current evidence and clinical importance. *Clin Res Cardiol* 2022; 111: 994–1009. 2022/03/17.
- Fleiss JL, Levin B and Paik MC. The measurement of interrater agreement. In: Shewart WA and Wilks SS (eds) *Statistical methods for rates and proportions*. Hoboken, NJ: John Wiley & Sons, 2003, pp.598–626.
- Bashar SK, Han D, Hajeb-Mohammadalipour S, et al. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. *Sci Rep* 2019; 9: 15054. 2019/10/23.
- Avram R, Ramsis M, Cristal AD, et al. Validation of an algorithm for continuous monitoring of atrial fibrillation using a consumer smartwatch. *Heart Rhythm* 2021; 18: 1482–1490. 2021/04/11.
- Zhu L, Nathan V, Kuang J, et al. Atrial fibrillation detection and atrial fibrillation burden estimation via wearables. *IEEE J Biomed Health Inform* 2022; 26: 2063–2074. 2021/12/03.
- Merschel S and Reinhardt L. Analyzability of photoplethysmographic smartwatch data by the preventicus heartbeats algorithm during everyday life: feasibility study. *JMIR Form Res* 2022; 6: e29479. 2022/03/29.
- Ford C, Xie CX, Low A, et al. Comparison of 2 smart watch algorithms for detection of atrial fibrillation and the benefit of clinician interpretation: SMART WARS study. *J Am Coll Cardiol EP* 2022; 8: 782–791. 2022/06/24.
- 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. 2018/04/18.
- Svennberg E, Tjong F, Goette A, et al. How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. *Europace* 2022; 24: 979–1005. 2022/04/04.
- Ishiguchi H, Shimizu A, Ishikura M, et al. Association between atrial high-rate episodes and ischemic/Major bleeding events in patients with a cardiac implantable electronic

- device- A 10-year, single-center historical cohort study. *Circ J* 2021; 85: 1329–1337. 2021/04/20.
21. Pastori D, Miyazawa K, Li Y, et al. Atrial high-rate episodes and risk of major adverse cardiovascular events in patients with cardiac implantable electronic devices. *Clin Res Cardiol* 2020; 109: 96–102. 2019/05/31.
  22. 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. 2017/03/23.
  23. Witt CT, Kronborg MB, Nohr EA, et al. Early detection of atrial high rate episodes predicts atrial fibrillation and thromboembolic events in patients with cardiac resynchronization therapy. *Heart Rhythm* 2015; 12: 2368–2375. 2015/07/15.
  24. Demkowicz PC, Dhruva SS, Spatz ES, et al. Physician responses to Apple watch-detected irregular rhythm alerts. *Am Heart J* 2023; 262: 29–37. 2023/04/22.
  25. Benjamin EJ, Go AS, Desvigne-Nickens P, et al. Research priorities in atrial fibrillation screening: a report from a national heart, lung, and blood institute virtual workshop. *Circulation* 2021; 143: 372–388. 2021/01/26.
  26. Staerk L, Sherer JA, Ko D, et al. Atrial fibrillation: epidemiology, pathophysiology, and clinical outcomes. *Circ Res* 2017; 120: 1501–1517. 2017/04/30.
-