ESC European Society of Cardiology

European Heart Journal - Digital Health (2023) **4**, 402–410
https://doi.org/10.1093/ehjdh/ztad039

ORIGINAL ARTICLE

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

Pamela Reissenberger [1], Peter Serfözö[1], Diana Piper[2], Norman Juchler[3], Sara Glanzmann[1], Jasmin Gram[1], Karina Hensler[1], Hannah Tonidandel[1], Elena Börlin[4], Marcus D'Souza[4,5], Patrick Badertscher [6], and Jens Eckstein[1,4,*]

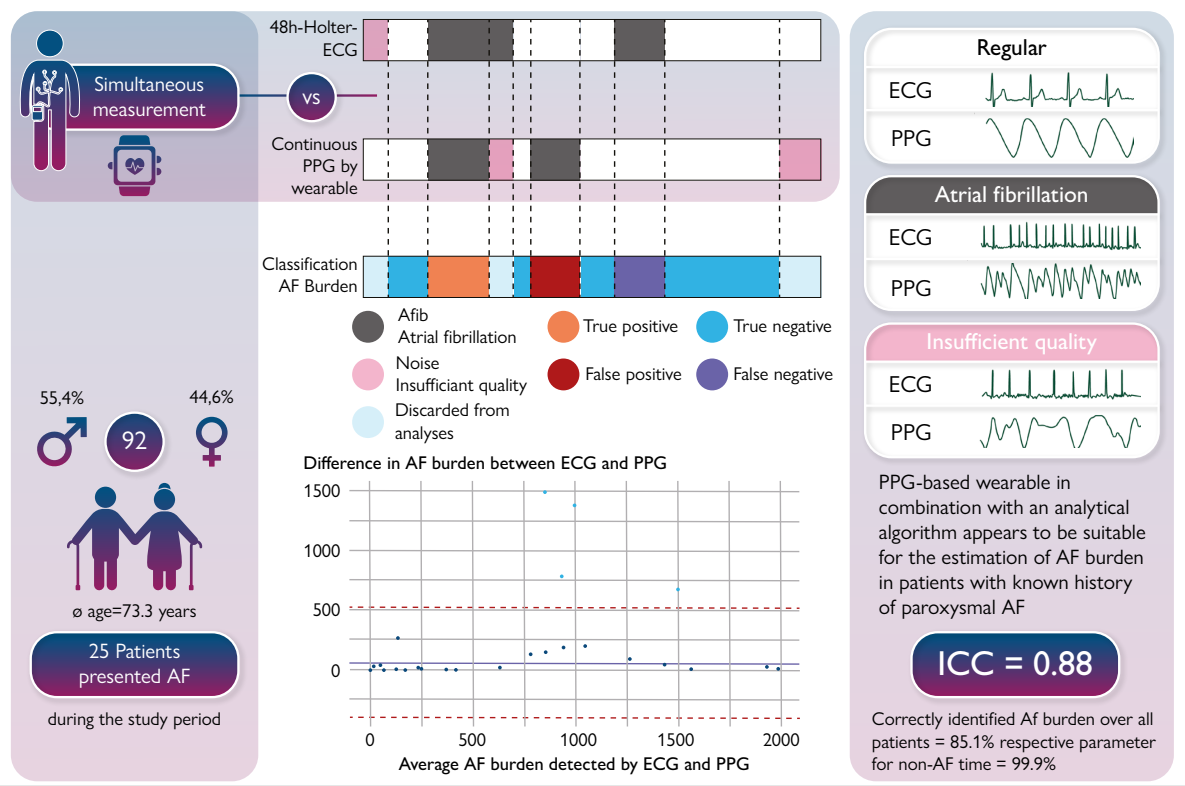
[1]Department of Internal Medicine, University Hospital Basel, Petersgraben 4, 4031 Basel, Switzerland; [2]Preventicus, Ernst-Abbe-Str. 15, 07743 Jena, Germany; [3]Institute of Computational Life Sciences, Zurich University of Applied Sciences, Schloss 1, 8820 Wädenswil, Switzerland; [4]Department Digitalization & ICT, University Hospital Basel, Spitalstrasse 26, 4031 Basel, Switzerland; [5]Department of Neurology, University Hospital Basel, Petersgraben 4, 4031 Basel, Switzerland; and [6]Department of Cardiology, University Hospital Basel, Petersgraben 4, 4031 Basel, Switzerland

Received 19 February 2023; revised 18 May 2023; online publish-ahead-of-print 6 July 2023

Aims	Recent studies suggest that atrial fibrillation (AF) burden (time AF is present) is an independent risk factor for stroke. The aim of this trial was to study the feasibility and accuracy to identify AF episodes and quantify AF burden in patients with a known history of paroxysmal AF with a photoplethysmography (PPG)-based wearable.
Methods and results	In this prospective, single-centre trial, the PPG-based estimation of AF burden was compared with measurements of a conventional 48 h Holter electrocardiogram (ECG), which served as the gold standard. An automated algorithm performed PPG analysis, while a cardiologist, blinded for the PPG data, analysed the ECG data. Detected episodes of AF measured by both methods were aligned timewise.Out of 100 patients recruited, 8 had to be excluded due to technical issues. Data from 92 patients were analysed [55.4% male; age 73.3 years (standard deviation, SD: 10.4)]. Twenty-five patients presented AF during the study period. The intraclass correlation coefficient of total AF burden minutes detected by the two measurement methods was 0.88. The percentage of correctly identified AF burden over all patients was 85.1% and the respective parameter for non-AF time was 99.9%.
Conclusion	Our results demonstrate that a PPG-based wearable in combination with an analytical algorithm appears to be suitable for a semiquantitative estimation of AF burden in patients with a known history of paroxysmal AF.
Trial Registration number	NCT04563572.

* Corresponding author. Tel: +41 61 3297689, Email: jens.eckstein@usb.ch

Graphical Abstract



Keywords Atrial fibrillation • AF burden • Photoplethysmography • Smartwatch • Monitoring

Introduction

Atrial fibrillation (AF) is related to a significant burden for global healthcare systems due to associated morbidity and mortality.[1] Present treatment recommendations for the prevention of cardioembolic stroke are based on the diagnosis of AF and vascular risk factors (implemented in the CHA2DS2-VASc score).[2] However, recent data from implantable cardiac devices and wearable sensors suggest that, for the determination of the individual stroke risk, AF might not be taken as a binary state (present or absent) but as a more differentiated factor.[3,4] The gradual impact of AF on stroke risk seems to correlate with AF burden, which can be defined as the amount of time an individual spends in AF.[2] Atrial fibrillation burden can be measured in a quantitative way if continuous monitoring devices are used. It can be defined by the number and longest duration of AF episodes during a monitoring period, or the proportion of time in AF during a monitoring period.[4] Clinical studies have evaluated the correlation between subclinical AF burden detected by implanted devices and stroke risk (ASSERT; TRENDS, SOS AF).[5–7] They found that AF episodes lasting longer than 24 h or AF burden >5.5 h/day were associated with an increased risk for stroke. Accordingly, recent studies suggest that a higher AF burden is associated with a higher risk for stroke, heart failure, and mortality.[3,8–10] Wrist-worn photoplethysmography (PPG)-based monitoring devices have been shown to provide high accuracy AF detection.[11–13] These devices represent a promising tool for low cost, convenient to wear, easily accessible, near continuous AF burden monitoring in a broad population.[14] In this clinical trial, our aim was to study the feasibility and accuracy of a non-invasive PPG-based monitoring device coupled with an AF screening algorithm to identify AF episodes and determine AF burden in patients with a known history of paroxysmal AF.

Methods

Participants

Ethics approval for this prospective, single-centre trial was obtained from the local ethics committees (BB 141/16, EKNZ BASEC 2016-01175). Ambulatory and hospitalized patients of the University Hospital Basel, Switzerland, were screened for inclusion and exclusion criteria on the basis of electronic patients' records. Patients 18 years old or older, with a known history of paroxysmal atrial fibrillation and no implanted cardiac device (e.g. pacemaker, implantable cardioverter defibrillator) were eligible for the study. If these conditions were met, patients were recruited after giving written informed consent.

Investigational products

The algorithm used in this trial was developed for the screening of AF episodes and quantification of AF burden based on a PPG signal (Heartbeats algorithm, Preventicus®, Jena, Germany, version 1.1.4). The algorithm is clinically validated, CE marked, and certified as medical device (Class IIa).[11,15] Study participants were equipped with a wrist-worn PPG sensor integrated in a smartwatch or a bracelet (CardioWatch 287-1, Corsano Health B.V., Bussum, The Netherlands, manufactured by MMT, Geneva,

Switzerland), a smartphone (Samsung Galaxy A40, Android OS, Samsung Electronics Co., Ltd, Seoul, South Korea) and a Holter ECG. The PPG sensor integrated in both devices has been certified as medical device under EU-MDR standards (European Union Medical Devices Regulations), bears CE conformity marking, and uses the sampling frequency of 25 Hz. The PPG sensor and the smartphone were linked via Bluetooth using an App provided by the sensor manufacturer (MMT-365-App, Geneva, Switzerland). Holter ECGs (SEER™ 1000, GE Healthcare, Getemed AG, Freiburg, Germany) served as gold standard to compare with the PPG signal-based data. The devices use a sampling frequency of 256 Hz and generate two leads (by five electrodes).

Data collection

During the per-protocol monitoring time of 48 h, a PPG and a Holter ECG measurement were obtained in parallel. The PPG sensor and the Holter ECG were synchronized at the beginning of the measurement. Patients were asked to document their symptoms or events during the monitoring period in a simplified Holter Monitor Diary. After the monitoring period, the patients returned all devices to the trial team, which downloaded all the data to a local database.

Processing of photoplethysmography recordings

De-identified PPG files were then analysed with the Heartbeats algorithm by company staff, who were blinded for all other data, including ECG data and medical information. The algorithm evaluates data segments that span 60 consecutive seconds (timestamp xx:00 to xx:59) using a complex non-linear combination analysis comprising beat-to-beat changes of pulse wave time intervals and pulse wave morphology to discriminate between SR and atrial arrhythmia. It assigns for each of these 1 min segments one of the three possible labels: 'Afib' (Atrial fibrillation), 'SR' (sinus rhythm), or 'Noise' (e.g. insufficient quality). If more than 10% of the 1 min segment in the PPG signal is disturbed by motion artefacts or other disruptive factors, the whole segment is considered 'Noise' by the algorithm and excluded from the analysis. Subsequently, technicians of an associated telecare provider (Telecare, Ulm, Germany) performed a visual quality control of suspected AF episodes in the PPG data. The technicians visually analyzed the PPG signal without having any access to the corresponding ECG data. If the technicians saw irregular heartbeats and a high variability in the heart rate, they confirmed the suspicious segment as AF episode. Each confirmed AF episode in the PPG data, labelled as 'Afib', and each 'Noise' episode, was documented with two timestamps marking start and end of the episodes.

Interpretation of electrocardiogram recordings

The ECG recordings were analysed using standard software (Cardioday® V2.5, GE Healthcare, Getemed AG, Freiburg, Germany). The first step of the ECG interpretation was an automated analysis by the ECG software. In a second step, the supervising physician examined the ECG data and evaluated if the automated interpretation was correct or had to be rectified. Then a senior cardiologist validated the ECG analysis and corrected if necessary. In a third step, a second senior cardiologist and electrophysiology fellow was consulted for remaining uncertainties and to validate corrections made after the automated analysis. Consensus was reached in all uncertainties. The clinical experts analysed the ECG data for 'Afib' episodes and 'Noise' episodes, which were documented with two timestamps marking start and end of the episodes. The monitoring time that was not labelled as 'Afib' or 'Noise' was automatically labelled as 'Normal Rhythm'. This included regular sinus rhythm (SR) and other rhythms like supraventricular tachycardia (including AV-node reentry tachycardia), atrial or ventricular premature beats, or sinus arrhythmia with irregular beats but with distinguishable P-waves. An AF episode was defined as absolute atrial arrhythmia with a length ≥1 min. This definition deviates from the more common definition of an absolute atrial arrhythmia lasting longer than 30 s,[2] because the PPG algorithm needs at least 1 min to decide whether the segment is 'Afib', 'SR', or 'Noise'. Atrial flutter was counted as AF, because according to the European Society of Cardiology (ESC) guidelines, it has the same therapeutic consequences as AF.[2] Contrary to the PPG data, an AF episode in the ECG data was annotated on the exact beat the arrhythmia began and therefore, the AF episode could begin on any given second of a minute. Disturbed data segments of insufficient quality, which took up more than 30 s at a stretch, were labelled as 'Noise' and excluded from the analysis. In 'Normal Rhythm' segments, the signal was labelled as 'Noise' if the P-wave was not discernible due to artefacts. 'Afib' episodes were labelled as interrupted by 'Noise' episodes if the beats were indistinguishable from each other and therefore, the arrhythmia was not recognizable. If more than one SR beat occurred subsequently during an 'Afib' episode, the episode was interrupted and divided into two. All physicians were blinded for the PPG analysis.

Data merging

After the ECG analysis was completed, results of the PPG analysis were returned to the University Hospital Basel, re-identified, and merged with the ECG diagnosis for statistical analysis. The data merging was performed using a script in R studio (Version 2022.02.3-492, R Studio, Inc., Boston, MA, USA), which was written in cooperation with the Preventicus Company and the supervising physician at the University Hospital Basel. An independent data scientist of the Zurich University of Applied Sciences checked the script for errors and possible biases. As an initial step, the R script aligned the PPG and ECG measurements timewise to assess the evaluable monitoring time (simultaneously monitored time without 'Noise' segments in either PPG or ECG signal, as seen in Supplementary material online, *Figure S1*). Then the 'Afib' episodes of the PPG data, as documented by the respective timestamps, were compared with those of the ECG data for each patient (as illustrated in *Figure 1*). The correct identification of an AF episode (true positive episode) was defined as an overlap of an AF episode detected by both PPG and ECG analyses lasting longer than 10 consecutive seconds. The detected AF burden was defined as the cumulative amount of time in AF and was assessed in seconds first and later converted to AF burden minutes for practical reasons. 'Afib' seconds that were detected by both methods were labelled true positive (yellow segment in *Figure 1*), those only detected by the ECG signal were labelled false negative (blue segment in *Figure 1*), and those that were only seen in the PPG data were labelled false positive (red segment in *Figure 1*). True negative non-AF time was defined as time that was neither labelled as 'Afib' or 'Noise' in both methods, shown as green segment in *Figure 1*. To compare patients with differing monitoring times, a percentage of time in AF (ratio between time in AF and evaluable monitoring time) was calculated for each patient.

Statistical analysis

Categorical variables are reported as absolute numbers and percentage and continuous variables are reported as median or mean with standard deviation and ranges. The statistical analysis was performed in R studio. To assess the diagnostic accuracy of the PPG-based estimation of AF burden compared with the ECG gold standard, a percentage of correctly identified AF burden minutes and a percentage of correctly identified non-AF time were calculated for each patient. The percentage was calculated by dividing the true positive or the true negative minutes by the amount of AF burden or non-AF time detected by the ECG. The proportion of correct detections over all patients was estimated using a logistic regression model. In this model, the correctness of the PPG signal for each 1 min time window was regressed on just an intercept plus a random effect for the individual patient. By doing so, the across-patient odds of correct detection are estimated by the intercept, while allowing each patient to have his/her deviation from these across-patient odds. Transforming the odds intercept to a proportion gives an estimate of the across-patients proportion of correct detections and allows us to correctly estimate this proportion's standard error and thus confidence interval. An intraclass correlation coefficient (ICC) analysis with a two-way mixed effect model was used to assess the agreement between the ECG and PPG measures concerning AF burden minutes, non-AF time, and percentage of time in AF. The agreement is illustrated in a Bland–Altman plot.

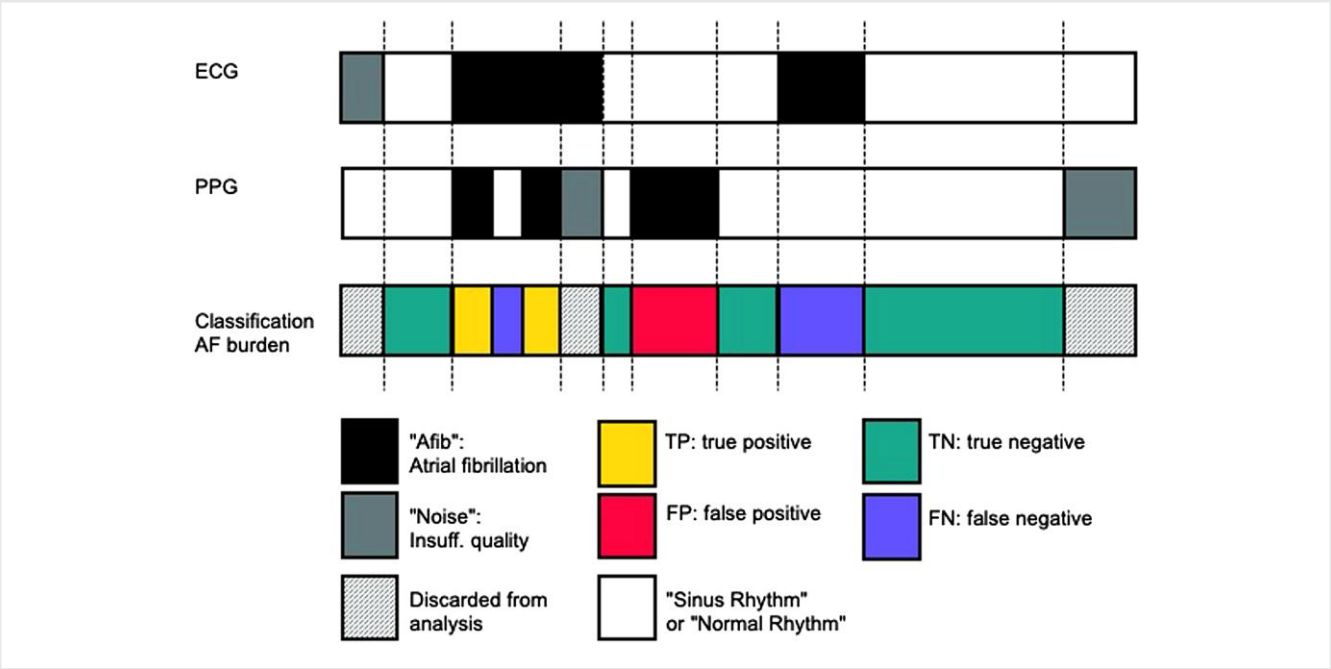


Figure 1 Conceptual illustration of atrial fibrillation burden classification per patient. First, the atrial fibrillation burden is assessed independently using photoplethysmography and electrocardiogram signals (upper and middle row). Using the electrocardiogram-based assessment as gold standard, the atrial fibrillation burden assessment is then classified into true/false positive/negative (bottom row). The time unit of this classification is 1 s.

Results

Patient characteristics

One hundred patients were enrolled between October 2020 and December 2021. Eight patients had to be excluded from the analysis. Four due to malfunction of the PPG measurement, two due to unreturned devices, and two due to impaired time matching between PPG and ECG signal. Therefore, a total of 92 patients were included in the analysis. Patient characteristics are listed in *Table 1*. Included were 41 females (44.6%) and 51 males (55.4%). The mean age was 73.3 (standard deviation, SD: 10.4) years. Prior to the study, all patients were diagnosed with paroxysmal AF. However, during the study period, 25 patients (27.2%) had AF episodes detected by the ECG in the evaluable monitoring time and 67 had no detected AF episodes. In order to simplify these findings in the following results, we divided the patients into a group that did not have any AF episodes during the study period (non-AF group) and a group that did have AF episodes (AF group). The AF group was older, 74.5 (SD: 9.8) vs. 72.8 (SD: 10.6), but the difference proved to be insignificant (P-value = 0.494).

Evaluable monitoring time

Due to technical issues with the PPG sensors and patient incompliance, the per-protocol simultaneous monitoring time of PPG and ECG measurements (48 h) was reduced to 43.7 h on average (2622 min), median 47.8 h (2870 min). For additional information on simultaneous monitoring time, see Supplementary material online, *Figure S3*. The percentage of signals classified as noise (proportion of time that had to be excluded from analysis) was, on average, 50.7% (SD: 20.5%) in the PPG method and 8.2% (SD: 12.6%) in the ECG method. For detailed information on noise percentage and evaluable monitoring time, see Supplementary material online, *Figures S2–S4* and *Table S3*.

Atrial fibrillation burden estimation performance

A total AF burden of 355.6 h (21 336 min) with an average of 14.2 h (853 min) per patient (SD: 686 min; range: 1–1991 min; median: 847 min) were detected by ECG in the AF group. The PPG algorithm labelled a total of 317.8 h (19 070 min) as AF suspicious. A total of 272.9 h (16 371 min) were confirmed as PPG AF episodes by technicians of an associated telecare provider and the remaining 45 h (2699 min) were not confirmed, either due to poor signal quality that was not automatically detected or other rhythm disorders which were misclassified by the algorithm as AF. After comparison with the ECG analysis, the PPG analysis correctly identified an AF burden (true positive AF burden) on total of 262.2 h (15 729 min) and an average of 10.5 h (629 min) per patient (SD: 610 min; range: 0–1982 min; median: 416 min). Three patients of the AF group did not have any true positive AF in the PPG measurements. A total of 10.7 h (642 min) in the PPG analysis were false positive, distributed in five patients of the non-AF group. There were no false positive AF burden minutes detected by the PPG in the AF group. Results of measured AF burden in minutes detected by the ECG, true positive AF burden in minutes detected by the PPG, the monitoring time in minutes, and the evaluable monitoring time for each patient in the AF group are displayed in *Table 2*. The agreement between the ECG and PPG measures concerning AF burden was assessed by calculating an ICC of 0.88 and is visually illustrated in a Bland–Altman plot (*Figure 2*). The ICC for agreement concerning non-AF time was 1, as illustrated in a Bland–Altman plot (*Figure 3*). If the time segments with disturbed signal in the PPG would not have been excluded from the ECG analysis, the ECG would have detected an AF burden on total of 764.2 h (45 849 min) in 26 patients and an average of 29.4 h (1763 min) per patient (SD: 1183 min; ranging from 1–2870 min; median: 2440 min). The difference between the total amount of AF burden detected by the ECG during the

Table 1 Patients characteristics over all patients and stratified by atrial fibrillation

Patient characteristics	N (%) or mean (SD)		
	Overall	AF group	Non-AF group
N	92	25	67
Age in years [mean (SD)]	73.26 (10.37)	74.48 (9.81)	72.81 (10.60)
Gender = male (%)	51 (55.4)	14 (56.0)	37 (55.2)
Setting = hospitalized (%)	53 (57.6)	15 (60.0)	38 (56.7)
Measurement device = smartwatch (%)	59 (64.1)	12 (48.0)	47 (70.1)
AF = Non-AF (%)	67 (72.8)		
BMI category (%)			
1 ≤ 18.5 (underweight)	4 (4.4)	1 (4.0)	3 (4.5)
2 = 18.5–24.9 (normal weight)	37 (40.7)	9 (36.0)	28 (42.4)
3 = 25–29.9 (overweight)	31 (34.1)	7 (28.0)	24 (36.4)
4 ≥ 30 (obese)	19 (20.9)	8 (32.0)	11 (16.7)
Skin colour Fitzpatrick scale (%)			
1	14 (15.2)	2 (8.0)	12 (17.9)
2	47 (51.1)	15 (60.0)	32 (47.8)
3	21 (22.8)	6 (24.0)	15 (22.4)
4	7 (7.6)	0 (0.0)	7 (10.4)
5	3 (3.3)	2 (8.0)	1 (1.5)
Tattoo in PPG sensor area = Yes (%)	1 (1.1)	1 (4.0)	0 (0.0)
Hairiness (%)			
1 = mild	60 (65.2)	14 (56.0)	46 (68.7)
2 = moderate	23 (25.0)	7 (28.0)	16 (23.9)
3 = excessive	9 (9.8)	4 (16.0)	5 (7.5)
CHA_2DS_2-VASc score (%)			
0	4 (4.3)	1 (4.0)	3 (4.5)
1	7 (7.6)	1 (4.0)	6 (9.0)
2	15 (16.3)	4 (16.0)	11 (16.4)
3	17 (18.5)	5 (20.0)	12 (17.9)
4	18 (19.6)	4 (16.0)	14 (20.9)
5	13 (14.1)	4 (16.0)	9 (13.4)
6	14 (15.2)	3 (12.0)	11 (16.4)
7	2 (2.2)	1 (4.0)	1 (1.5)
8	2 (2.2)	2 (8.0)	0 (0.0)

AF, atrial fibrillation.
The described BMI categories were defined by the World Health Organization (WHO).
Skin colour was assessed using the Fitzpatrick scale and hairiness of the skin in the PPG sensor area was assessed using a subjective visual rating scale.
The CHA2DS2-VASc Score is a clinical score to assess stroke risk for patients with atrial fibrillation based on gender, age, and vascular risk factors.

evaluable monitoring time and the AF burden that was detected by the ECG during the simultaneous monitoring time was referred to as 'missed' AF burden. An overview of the mentioned results is displayed in *Table 3*. The percentage of correctly identified AF burden over all patients as calculated by using the described logistic mixed regression model was 85.1% (95% confidence interval 57.9–95.8%) and the percentage of correctly identified non-AF time over all patients was 99.9% (95% confidence interval 99.9–100.0%). The ICC concerning the percentage of time in AF measured by the ECG and by the PPG was 0.91. The agreement between the two measurement methods is illustrated in a Bland–Altman plot (*Figure 4*).

Atrial fibrillation episode detection by PPG was distorted because of fragmentation of PPG episodes due to disturbed data segments. See Supplementary material online, *Figures S5–S7* for AF episode detection analysis and Supplementary material online, *Figure S8* and *Table S1, S2*, and *S4* for sub-group analysis of ambulatory and hospitalized patients.

Discussion

This study evaluated the accuracy of a wrist-worn PPG-based monitoring device coupled with an AF screening algorithm to identify AF episodes and determine AF burden. Due to fragmentation by noisy signals, the absolute quantification of AF episodes did not prove to be a suitable parameter to quantify AF burden. It would have been possible to summarize multiple PPG AF episodes overlapping with a single ECG AF episode and count them as a single AF episode as described earlier by Wasserlauf *et al.*[12] This, however, would have resulted in the overestimation of AF episode detection accuracy. True positive AF burden in minutes per patient, measured by PPG, showed a high correlation to the respective parameter measured by the Holter ECG as visualized in the Bland–Altman plot (*Figure 2*) and the ICC. Although the PPG measurement was prone to underestimate the AF burden per patient, the difference between the two measurement methods was evenly distributed among patients with lower AF burden and higher AF burden (*Figure 2*). The percentage of correctly identified AF burden by PPG in relation to the Holter ECG was high (85.1%). Five patients, however, had a particularly low percentage, which resulted in a wide limit of agreement of ±500 min. Three of the 25 patients in the AF group had zero detected AF burden minutes in the PPG analysis. Analysing patient characteristics that could lead to a potential decrease in PPG sensitivity like skin tone, tattoos in the PPG sensor area, hairiness[16] or the BMI (due to different optical properties of the skin in obese patients[17,18]) did not lead to any possible explanation for the reduced percentage of correctly identified AF burden in these patients. In general, our patient collective consisted of a heterogeneous group concerning the properties skin tone, hairiness and BMI, as shown in *Table 1*. Analysing the original ECG data of these patients showed that two of them presented a frequent alternation between atrial fibrillation and atrial flutter. The third patient only had 1 min of AF detected by ECG and the other two patients did not show any exceptional rhythm pattern. Atrial flutter with a relatively regular heart rate can lead to AF detection problems by PPG, because there is little to no variation between the beats, leading to a higher rate of false negative AF burden minutes. The amount of false positive AF burden was low, described as a high percentage of correctly identified non-AF time. False positive AF burden minutes were detected in 5 of 67 patients in the non-AF group and correlated with a high rate of irregular beats in the ECG signal due to atrial and ventricular premature beats. Therefore, the PPG signal did not overestimate the AF burden in patients with true AF episodes during the study period. In order to generalize the findings for patients with differing monitoring times due to technical issues, we assessed the relative parameter of percentage of time in AF per patient and per measurement method. The correlation of the two measurement methods, as described by the ICC and visualized in the Bland–Altman plot (*Figure 4*), was high. Due to the reason that many patients (13 out of 25) had a percentage of time in AF of 100% measured by the gold standard, the average measurement

Table 2 Results of measured atrial fibrillation burden in minutes detected by the electrocardiogram, true positive atrial fibrillation burden in minutes detected by the photoplethysmography (with the corresponding percentage of true positive atrial fibrillation burden measured by the photoplethysmography in comparison to atrial fibrillation burden measured by electrocardiogram), the monitoring time in minutes, and the evaluable monitoring time in minutes (with the corresponding percentage of monitoring coverage) for each patient in the atrial fibrillation group

ID AF group	AF burden minutes measured by the ECG	True positive AF burden minutes measured by the PPG (percentage of true positive AF burden measured by the PPG in comparison to AF burden measured by ECG)	Monitoring time in minutes	Evaluable monitoring time in minutes (percentage of monitoring coverage)
3	68	27 (40%)	2575	609 (24%)
5	268	0 (0%)	2871	1656 (58%)
8	1837	1160 (63%)	2870	1866 (65%)
9	65	65 (100%)	2870	1693 (59%)
10	1324	538 (41%)	2871	1324 (46%)
12	1945	1918 (99%)	2870	1945 (68%)
20	1566	1558 (99%)	2870	1566 (55%)
22	130	122 (94%)	2340	130 (6%)
24	1309	1216 (93%)	2870	1310 (46%)
30	1991	1981 (99%)	2869	1991 (69%)
32	1685	304 (18%)	2868	1685 (59%)
33	929	779 (84%)	2869	929 (32%)
34	1146	944 (82%)	2541	1146 (45%)
39	641	620 (97%)	2870	1436 (50%)
41	1035	846 (82%)	2870	1715 (60%)
42	417	416 (100%)	2862	1677 (59%)
48	253	242 (96%)	2833	2175 (77%)
56	847	714 (84%)	2868	847 (30%)
63	1	0 (0%)	2871	1410 (49%)
75	1597	104 (7%)	2870	1597 (56%)
76	243	223 (92%)	2864	1346 (47%)
86	372	367 (99%)	2870	372 (13%)
95	1456	1409 (97%)	2869	1456 (51%)
98	31	0 (0%)	1314	759 (58%)
100	171	170 (99%)	511	171 (33%)

between the two methods increased in a linear fashion in relation to a decrease in the difference of the two measurements. This explains the linear relation of the dots on the right-hand site of the Bland–Altman plot (*Figure 4*).

The PPG method was prone to artefacts in the measurement data leading to a high noise percentage and thereby reducing the evaluable monitoring time. This can result in an uncertainty concerning the true amount of AF burden of the individual patient during the monitoring period due to missed AF episodes. It could especially be a problem in out-patients, who are an important target population for continuous AF monitoring in the future, because they would be especially prone to a high amount of motion artefacts during everyday life activities. We analysed the 'missed' AF burden, describing the amount of AF burden detected by the ECG in the simultaneous monitoring time but undetected by the PPG signal, because of noise. The total of 'missed' AF burden was approximately the same amount as the detected AF burden, which correlates to the noise percentage of 50.7%. However, the noise percentage of the PPG method in our trial was comparable to previous studies with similar conditions. Wasserlauf *et al.*[12], for example, were able to analyse, on average per patient, 11.3 h of data from a 24 h monitoring period, which would translate to 53% of noise and Chang *et al.*[19] were able to analyse 7.7 h, on average per patient, out of a 24 h monitoring period, which would translate to 68% of noise.

There are several possibilities to reduce the noise percentage of the PPG measurement data. Earlier studies investigating the detection performance of AF by PPG-based wearables took the measurements for only a short monitoring time and in resting positions leading to evidence mostly based on motionless patients and thus keeping the signal disruption by motion artefacts low. This approach, however, is not suitable to assess the continuous, long-term AF detection performance in daily life necessary to estimate AF burden. Another possibility is to experiment with different sensors and bracelets of the monitoring devices. Individual bracelet materials and lengths suited to the individual participant's anatomy could improve the sensor-to-skin contact and therefore reduce artefacts. A different approach would be to increase the robustness of the PPG algorithm towards disturbed signals so that the algorithm would be able to monitor AF even during motion artefact periods. Zhu *et al.*[20] for example, were able to reduce the noise percentage to 32.2% during a 28-day monitoring period by using innovative algorithm blocks and system designs. And lastly an alternative would be to study the relationship between noise percentage and the amount of 'missed' AF burden and to extrapolate the correctly identified AF burden measured by the PPG closer to the true AF burden.

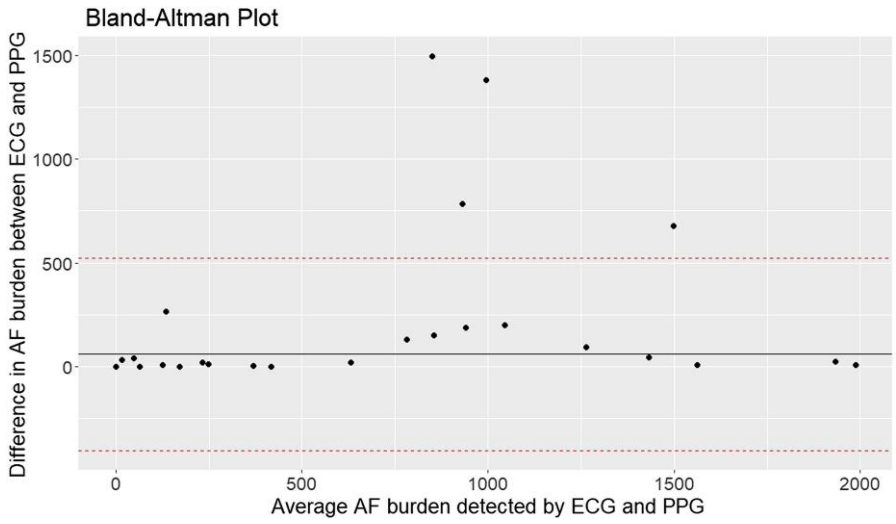


Figure 2 Bland–Altman analysis shows the agreement of atrial fibrillation burden in minutes per patient measured by photoplethysmography and electrocardiogram. Each dot represents the atrial fibrillation burden for each patient in the atrial fibrillation group. On the *x*-axis of the plot, the mean atrial fibrillation burden in minutes of the two measurement methods is displayed. On the *y*-axis, the difference between the atrial fibrillation burden in minutes measured by electrocardiogram and photoplethysmography is displayed. The true positive atrial fibrillation burden measured by the photoplethysmography was subtracted from the atrial fibrillation burden measured by the electrocardiogram. A positive difference between the two measurements is therefore resulting from a higher atrial fibrillation burden measured by the electrocardiogram than measured by the photoplethysmography. The black line represents the average value of difference between the atrial fibrillation burden in minutes measured by the electrocardiogram and the PPG. The red dashed lines represent the upper and lower 95% confidence interval of the difference between the measurement methods.

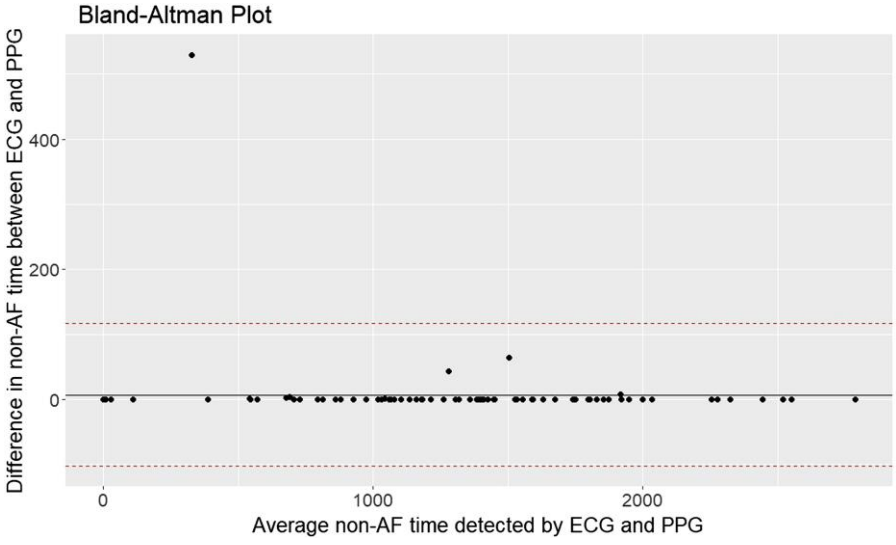


Figure 3 Bland–Altman analysis shows the agreement of non-atrial fibrillation time in minutes per patient measured by photoplethysmography and electrocardiogram. Each dot represents the non-atrial fibrillation time for each patient with non-atrial fibrillation time. On the *x*-axis of the plot, the mean non-atrial fibrillation time in minutes of the two measurement methods is displayed. On the *y*-axis, the difference between the non-atrial fibrillation time in minutes measured by electrocardiogram and photoplethysmography is displayed. The true negative non-atrial fibrillation time measured by the photoplethysmography was subtracted from the non-AF time measured by the electrocardiogram. A positive difference between the two measurements is therefore resulting from a higher non-atrial fibrillation time measured by the electrocardiogram than measured by the photoplethysmography. The black continuous line represents the average value of difference between the non-atrial fibrillation time in minutes measured by the electrocardiogram and the photoplethysmography. The red dashed lines represent the upper and lower 95% confidence interval of the difference between the measurement methods.

Table 3 Overall comparison of measured atrial fibrillation burden

	Total AF burden measured by ECG	AF-suspicious time measured by PPG	Telecare AF confirmed time measured by PPG	True positive AF burden measured by PPG	False positive AF burden measured by PPG	'Missed' AF burden measured by ECG outside of evaluable monitoring time
Total amount in minutes	21 336	19 070	16 371	15 729	642	24 513
Total amount in hours	355.6	317.8	272.9	262.2	10.7	408.6 (in 26 patients)

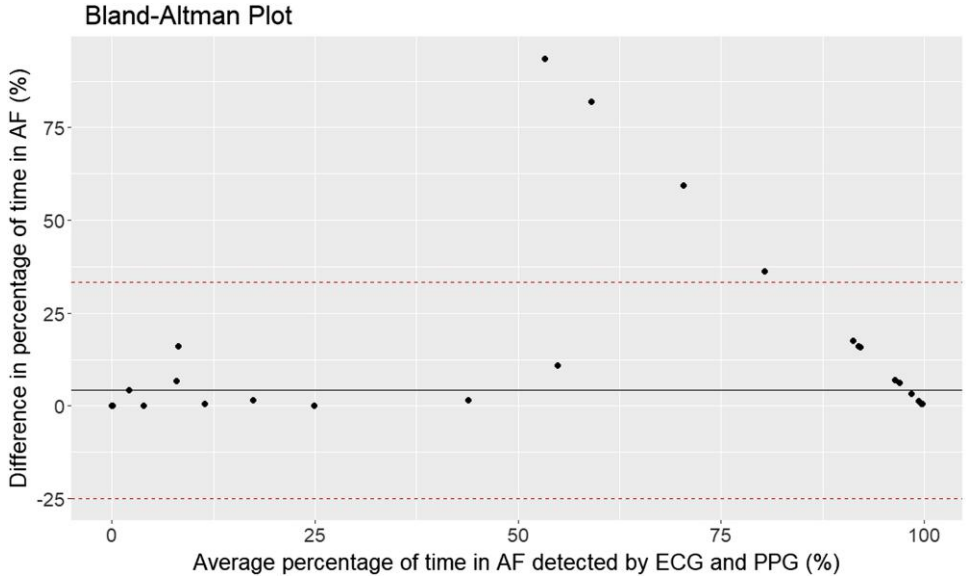


Figure 4 Bland–Altman analysis shows the agreement of percentage of time in atrial fibrillation per patient measured by photoplethysmography and electrocardiogram. Each dot represents the percentage of time in atrial fibrillation for each patient in the atrial fibrillation group. The percentage of time in atrial fibrillation per patient was calculated by dividing the individual amount of detected atrial fibrillation burden in minutes in the electrocardiogram for each patient and accordingly correctly identified atrial fibrillation burden in the PPG by the evaluable monitoring time for the respective patient. On the *x*-axis of the plot, the mean percentage of time in atrial fibrillation in minutes of the two measurement methods is displayed. On the *y*-axis, the difference between the percentage of time in atrial fibrillation measured by electrocardiogram and photoplethysmography is displayed. The black line represents the average value of difference between the percentage of time in atrial fibrillation measured by the electrocardiogram and the photoplethysmography. The red dashed lines represent the upper and lower 95% confidence interval of the difference between the measurement methods.

Our study had several limitations. Initially, we expected 75% of patients to have at least one AF episode detected by the ECG during the trial period because of the known history of paroxysmal AF. We expected to detect several AF episodes per patient during the monitoring time of 48 h. Ultimately, only 27% developed detectable AF episodes leading to a much smaller sample size of the AF group. This could indicate that a monitoring period of 48 h is too short. A larger patient cohort or a longer monitoring period could have compensated the lower prevalence of AF episodes. The inclusion of atrial flutter in the AF group could have led to a smaller percentage of correctly identified AF burden, because of the regular rhythm of atrial flutter, making it challenging for the algorithm to detect.[19] Nevertheless, it was included because of the same therapeutic consequences.[2] Another limitation was the reduction of evaluable monitoring time due to the exclusion of noise segments, which can result in an uncertainty of the true amount of AF burden of the individual patient. Experimenting with different PPG sensors and bracelets of the wrist-worn devices as well as increasing the robustness of the PPG algorithm towards disturbed signals could lead to a higher diagnostic yield.

Conclusion

Our results indicate that a PPG-based wearable in combination with an analytical algorithm can be used for a semiquantitative estimation of AF burden in patients with known history of paroxysmal AF. It has to be taken into account that due to the exclusion of noise segments the total evaluable monitoring time is reduced. Further refinement of the technology or noise robust algorithms could enhance the precision of AF burden quantification. With easily accessible and comfortable tools for AF burden estimation, such as PPG-based wearables, large-scale randomized controlled trials could be realized in order to identify the relation between AF burden and stroke risk to guide future recommendations for oral anticoagulation initiation.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

We would like to acknowledge the study support by Aura Winterhalder as the main study coordinator.

Funding

The study was funded by the European Commission for Research and Innovation (Eurostars: E! 12427—SmartAF) and the German Federal Ministry of Education and Research (BMBF, FKZ:01QE1859A).

Conflict of interest: D.P. is an employee of Preventicus. J.E. owns 0.5% virtual shares of Preventicus. All remaining authors have no conflicts of interest to declare.

Consent

All recruited patients gave written informed consent.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, *et al.* Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* 2014;**129**:837–847.
2. Hindricks G, Potpara T, Dagres N, Arbelo E, Boriani G, Castella M, *et al.* 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS): supplementary data. *Eur Heart J* 2021;**42**:373–498.
3. Chen LY, Chung MK, Allen LA, Ezekowitz M, Furie KL, McCabe P, *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. Potpara TS, Lip GYH, Blomstrom-Lundqvist C, Boriani G, Van Gelder IC, Heidbuchel H, *et al.* The 4S-AF scheme (stroke risk; symptoms; severity of burden; substrate): a novel approach to in-depth characterization (rather than classification) of atrial fibrillation. *Thromb Haemost* 2021;**121**:270–278.
5. Van Gelder IC, Healey JS, Crijns HJGM, Wang J, Hohnloser SH, Gold MR, *et al.* Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. *Eur Heart J* 2017;**38**:1339–1344.
6. Glotzer TV, Daoud EG, Wyse DG, Singer DE, Ezekowitz MD, Hilker C, *et al.* The relationship between daily atrial tachyarrhythmia burden from implantable device diagnostics and stroke risk: the TRENDS study. *Circ Arrhythm Electrophysiol* 2009;**2**:474–480.
7. Boriani G, Glotzer TV, Santini M, West TM, De Melis M, Sepsi M, *et al.* Device-detected atrial fibrillation and risk for stroke: an analysis of >10,000 patients from the SOS AF project (Stroke preventiOn Strategies based on Atrial Fibrillation information from implanted devices). *Eur Heart J* 2014;**35**:508–516.
8. Pandey A, Kim S, Moore C, Thomas L, Gersh B, Allen LA, *et al.* Predictors and prognostic implications of incident heart failure in patients with prevalent atrial fibrillation. *JACC Heart Fail* 2017;**5**:44–52.
9. Go AS, Reynolds K, Yang J, Gupta N, Lenane J, Sung SH, *et al.* Association of burden of atrial fibrillation with risk of ischemic stroke in adults with paroxysmal atrial fibrillation: the KP-RHYTHM study. *JAMA Cardiol* 2018;**3**:601–608.
10. Piccini JP, Passman R, Turakhia M, Connolly AT, Nabutovsky Y, Varma N. Atrial fibrillation burden, progression, and the risk of death: a case-crossover analysis in patients with cardiac implantable electronic devices. *Europace* 2019;**21**:404–413.
11. Dörr M, Nohturfft V, Brasier N, Bosshard E, Djurdjevic A, Gross S, *et al.* The WATCH AF trial: SmartWATCHes for detection of atrial fibrillation. *JACC Clin Electrophysiol* 2019;**5**:199–208.
12. 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.
13. Bonomi AG, Schipper F, Eerikäinen LM, Margarito J, van Dinther R, Muesch G, *et al.* Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photoplethysmography at the wrist. *J Am Heart Assoc* 2018;**7**:e009351.
14. Svennberg E, Tjong F, Goette A, Akoum N, Di Biase L, Bordachar P, *et al.* How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. *Europace* 2022;**24**:979–1005.
15. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S, *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.
16. Bousefsaf F, Maaoui C, Pruski A. Automatic selection of webcam photoplethysmographic pixels based on lightness criteria photoplethysmographic pixels selection based on lightness segmentation. *J Med Biol Eng* 2017;**37**:374–385.
17. Boonya-ananta T, Rodriguez AJ, Ajmal A, Du Le VN, Hansen AK, Hutcheson JD, *et al.* Synthetic photoplethysmography (PPG) of the radial artery through parallelized Monte Carlo and its correlation to body mass index (BMI). *Sci Rep* 2021;**11**:2570.
18. Yosipovitch G, DeVore A, Dawn A. Obesity and the skin: skin physiology and skin manifestations of obesity. *J Am Acad Dermatol* 2007;**56**:901–916; quiz 917–920.
19. Chang P-C, Wen M-S, Chou C-C, Wang C-C, Hung K-C. Atrial fibrillation detection using ambulatory smartwatch photoplethysmography and validation with simultaneous holter recording. *Am Heart J* 2022;**247**:55–62.
20. Zhu L, Nathan V, Kuang J, Kim J, Avram R, Olgin J, *et al.* Atrial fibrillation detection and atrial fibrillation burden estimation via wearables. *IEEE J Biomed Health Inform* 2022;**26**:2063–2074.