Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist

Alberto G. Bonomi, PhD; Fons Schipper, MSc; Linda M. Eerikäinen, MSc; Jenny Margarito, MSc; Ralph van Dinther, PhD; Guido Muesch, MSc; Helma M. de Morree, PhD; Ronald M. Aarts, PhD; Saeed Babaeizadeh, PhD; David D. McManus, MD; Lukas R.C. Dekker, MD

Background—Long-term continuous cardiac monitoring would aid in the early diagnosis and management of atrial fibrillation (AF). This study examined the accuracy of a novel approach for AF detection using photo-plethysmography signals measured from a wrist-based wearable device.

Methods and Results—ECG and contemporaneous pulse data were collected from 2 cohorts of AF patients: AF patients (n=20) undergoing electrical cardioversion (ECV) and AF patients (n=40) that were prescribed for 24 hours ECG Holter in outpatient settings (HOL). Photo-plethysmography and acceleration data were collected at the wrist and processed to determine the inter-pulse interval and discard inter-pulse intervals in presence of motion artifacts. A Markov model was deployed to assess the probability of AF given irregular pattern in inter-pulse interval sequences. The AF detection algorithm was evaluated against clinical rhythm annotations of AF based on ECG interpretation. Photo-plethysmography recordings from apparently healthy volunteers (n=120) were used to establish the false positive AF detection rate of the algorithm. A total of 42 and 855 hours (AF: 21 and 323 hours) of photo-plethysmography data were recorded in the ECV and HOL cohorts, respectively. AF was detected with >96% accuracy (ECV, sensitivity=97%; HOL, sensitivity=93%; both with specificity=100%). Because of motion artifacts, the algorithm did not provide AF classification for 44±16% of the monitoring period in the HOL group. In healthy controls, the algorithm demonstrated a <0.2% false positive AF detection rate.

Conclusions—A novel AF detection algorithm using pulse data from a wrist-wearable device can accurately discriminate rhythm irregularities caused by AF from normal rhythm. (*J Am Heart Assoc*. 2018;7:e009351. DOI: 10.1161/JAHA.118.009351.)

Key Words: arrhythmia (heart rhythm disorders) • cardioversion • screening • self-management

Atrial fibrillation (AF) is the most common cardiac arrhythmia, affecting around 3% of adults and 10% of those aged ≥75 years.[1,2] AF impairs patients' quality of life[3] and is associated with significant mortality,[4] increasing risk for stroke,[5] heart failure,[6] and hospitalization.[7,8] Early diagnosis of AF is critical to mitigate risk for stroke using oral anticoagulation therapy,[9–11] and monitoring AF is helpful to optimizing and evaluating the success of rhythm control strategies, including antiarrhythmic drug, cardioversion, and ablative therapies.[12] However, AF can be minimally or entirely asymptomatic and paroxysms of arrhythmia, owing to their sometimes brief nature, can elude clinical detection.[9] Timely and accurate detection of AF is therefore challenging.[13]

Conventional tools used to establish the presence of AF include 12-lead ECG or 24-hour Holter monitoring.[14] These established methods are effective to diagnose patients suffering from persistent forms of AF. However, many cases of paroxysmal AF remain undetected.[15] Data from patients with implantable cardiac devices demonstrate that intermittent, short (<30 days), or symptom-driven ECG recordings miss AF episodes with greater frequency (31–71% detection) as compared to more intensive monitoring.[16] Longer-term

---

Clinical Perspective

What Is New?

- A wrist-wearable device using photo-plethysmography can detect pulse irregularities attributable to atrial fibrillation in both hospital and home settings.

What Are the Clinical Implications?

- This technology represents an exciting tool that may be useful for long-term noninvasive monitoring of atrial fibrillation.
- Studies are needed to evaluate the performance of the technology in at-risk patient populations, guide optimization of the device from both patient and clinician perspectives, and guide implementation of this technology into health care for atrial fibrillation management and stroke prevention.

---

cardiac monitoring strategies (ie, implantable cardiac loop recorders) are costly and/or burdensome. Unobtrusive and inexpensive solutions to detect AF would greatly contribute to prevention of AF-related complications, including stroke and heart failure.[17]

In this study, we sought to investigate the performance of a novel AF detection strategy using pulse data from a wrist-based photo-plethysmography (PPG) recording device. This technology offers a low-cost, unobtrusive, and convenient solution for long-term patient monitoring. The objective was 2-fold: (1) to determine the performance of the algorithm for AF detection in a population of patients with known AF before and after electrical cardioversion and (2) to assess the AF detection accuracy among patients with AF as well as healthy adults.

Methods

Study Sample

A clinical trial was conducted to collect wrist PPG and ECG reference data in 2 different cohorts of patients: (1) 20 AF patients before and after elective electrical cardioversion (ECV) and (2) 40 patients prescribed to undergo 24/48-hour ECG Holter registration in daily living conditions (HOL) who had been diagnosed with paroxysmal or persistent AF. Patients were excluded from participation if they showed presence of any pathological cardiac arrhythmia different from AF. The study was approved by the medical ethical committee of MEC-U (Medical Research Ethics Committees United) in The Netherlands. All participants provided written informed consent. The study was registered with number: NL53827.100.15 at www.toetsingonline.nl. Data, analytical methods, and study materials will not be made available to

other researchers for purposes of reproducing the results or replicating the procedure.

Protocol

Pre- and post-cardioversion

Patients admitted for electrical cardioversion for AF were approached for informed consent. Consenting participants were asked to wear a wrist-based PPG measurement device (CM3 Generation-3, Connected Sensing; Philips, Eindhoven, The Netherlands) and a single-lead ECG apparatus (Actiwave Cardio, CamNtech Ltd, Cambridge, UK) before sedation, while lying in bed before the cardioversion procedure. Approximately 60 minutes of pulse and ECG data were acquired before and after the electrical cardioversion from all participants. Before delivery of the electric shock, the ECG measurement device was detached from the patient chest and reattached when the doctor confirmed that the procedure was finished. Cardioversion was successful in restoring normal sinus rhythm (NSR) in all patients. This protocol allowed gathering data during both AF and normal sinus rhythm under controlled, resting conditions.

Holter test

Patients referred for Holter monitoring as part of their usual clinical care were approached from a single center. After obtaining informed consent and after placement of a 12-lead ECG Holter monitor (H12+; Mortara, Milwaukee, WI, USA), the wrist-wearable PPG monitor was positioned on the participant's non-dominant arm. Participants were asked to wear the PPG monitor at home for the duration of the Holter monitoring period. When the recording period was complete, participants were asked to return the Holter and the wrist-wearable PPG device to the hospital.

Additional Data Sets

Two other data sets were used in this study only for validation purposes of the AF detection algorithm. An overview of all the data sets available for analysis is presented in Table 1.

Healthy control data set

A data set of wrist PPG waveforms recorded in daily living conditions was used to qualitatively assess the likelihood of false positive AF detection in a cohort of apparently healthy subjects. Volunteers were healthy adults recruited from 5 different observational studies. As part of the inclusion criteria, trial participants had to report no history of cardiac disease, including AF. These trials were approved by the Philips internal committee for biomedical experiments, and all participants provided written informed consent that allowed

Table 1. Overview of the Observational Trials and Data Sets Analyzed in This Study

Trial / Data Set	Inclusion Criteria	Dropout (No. of Patients)	Recording Duration	Measurement Condition	Types of Data	Reference	Analytical Purpose
ECV (n=20)	AF patients admitted for electrical cardioversion	Atrial flutter (2)	1 h before and 1 h after cardioversion	In-hospital bed rest	Wrist PPG ECG	Visual inspection of single-lead ECG	Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm
HOL (n=40)	Patients diagnosed with paroxysmal or persistent AF prescribed for ECG Holter	Atrial flutter (5) Atrial tachycardia (1)	24 to 48 h	Daily life	Wrist PPG ECG	Visual inspection of 12-lead ECG	Fine-tuning (50%) and holdout testing (50%) of the AF detection algorithm
Healthy Control I to V (n=120)							Assess the likelihood of false-positive AF detection
Trial I (n=18)	18 to 65 y, M/F healthy adults	None	3 times 24 h	Daily life	Wrist PPG	Not available	
Trial II (n=17)	40 to 65 y, M/F healthy adults	None	2 times 8 to 10 h	Bedtime monitoring	Wrist PPG	Not available	
Trial III (n=25)	18 to 65 y, M/F healthy adults	None	24 h	Daily life	Wrist PPG	Not available	
Trail IV (n=46)	18 to 65 y, M/F healthy adults working in clinical environment	None	3 times 24 h	Daily life (home and work shift)	Wrist PPG	Not available	
Trial V (n=14)	18 to 65 y, M/F professional truck drivers	None	5 to 7 d repeated up to 3 times	Daily life (home and on-road shift)	Wrist PPG	Not available	
MIT-BIH AF (n=25)	Patients with AF mostly paroxysmal	AF event <1 min (1)	10 h	Daily life	ECG	Visual inspection of ECG	Accuracy assessment for transient AF events

AF indicates atrial fibrillation; ECV, electrical cardioversion; HOL, Holter cohort; MIT-BIH AF, Massachusetts Institute of Technology-Beth Israel Hospital Atrial Fibrillation Database; M/F, male/female; PPG, photo-plethysmography.

retrospective use of the collected data for the purpose of this study. Participants were monitored using the wrist-based wearable device over a minimum of 7 hours and up to 4 weeks.

MIT-BIH AF data set

The MIT-BIH AF (Massachusetts Institute of Technology-Beth Israel Hospital Atrial Fibrillation Database) data set available at PhysioNet consists of 25 long-term ECG recordings of ≈10 hours in duration of patients suffering from paroxysmal AF (www.physionet.org).[18] This data set is commonly used for validation of ECG-based AF detection algorithms because it has a variety of AF cases. In our study, this data set was used to evaluate the ability of the AF detection algorithm to process heartbeat sequences and detect transient episodes of AF. One record (5091) was excluded from our study

because AF was only present for less than 1 minute in duration.

Apparatus

Wrist-wearable PPG and acceleration sensors

Cardiac rhythm and body motion were measured using a novel ambulatory monitoring modality embodied in a wrist-wearable device (Figure 1) equipped with the Cardio and Motion Monitoring Module (CM3 Generation-3), which included a PPG sensor and an accelerometer. The PPG sensor used a reflective modality, and the light source was formed by 2 green light LEDs. Sampling frequency of the raw data was 128 Hz, and the dynamic range of the accelerometer was ±8G. Validation of this PPG and accelerometer system to measure heart rate during sport and
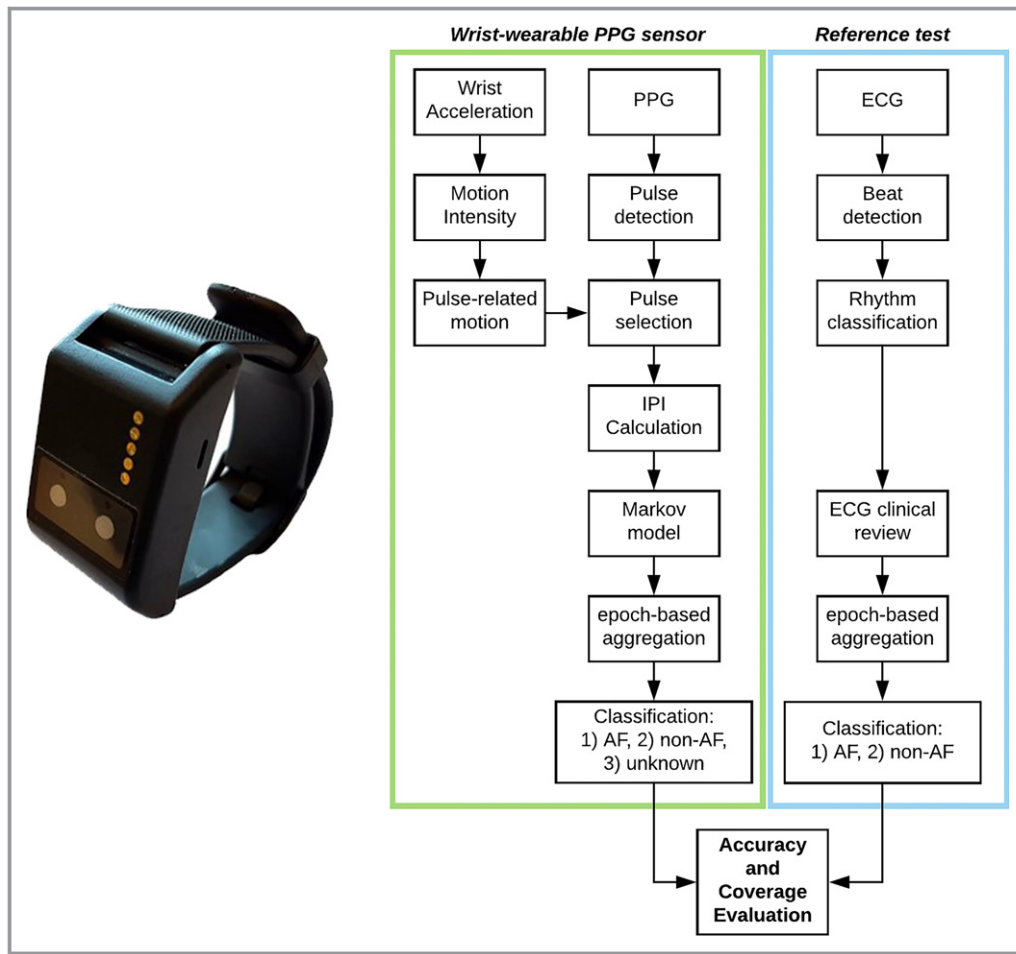
Figure 1. Image of the wrist-wearable sensor used to record photo-plethysmography (PPG) data in the clinical and observational trials (left); data processing flow chart for both the PPG and ECG sensors (right). Output of the PPG-based classification algorithm was either atrial fibrillation (AF), non-AF or unknown rhythm. IPI indicates inter-pulse interval.

activities of daily living was provided in previous publications.[19–21] The PPG sensor was able to sense changes in the reflected light from the skin caused by blood volume variations directly related to the cardiac systolic and diastolic cycle.

Data Processing
Inter-pulse intervals

The PPG waveform acquired at the wrist of patients was processed to: (1) determine the temporal location of fiducial points which corresponded to each heartbeat, according to a previously published method,[22] and (2) calculate the time interval between successive heart beats, here called the inter-pulse interval (IPI). The wrist acceleration signal was used to compute the amount of motion associated with each IPI. In case the motion intensity associated with a certain IPI exceeded a predefined threshold, such IPI was labeled as invalid and discarded from the analysis.[23]

AF detection algorithm

A first-order 11-states Markov model was used to calculate probability of AF given a sequence of IPI. Such an algorithm, inspired by the work of Moody and Mark,[24] has been previously trained and validated for AF detection on ECG data from Holter recordings.[25] Output from the Markov model was an AF probability reflecting the relative likelihood of observing a sequence of pulse intervals during AF versus making the same observation outside AF. In our study, the Markov model was adapted in order to process IPI sequences from PPG for AF detection. Such adaptation focused on fine-tuning the probability threshold applied to the Markov model output to identify episodes of AF (see Data S1 for details). Invalid IPIs, such as those obtained in the presence of considerable wrist motion, were not used to assess and update the AF probability. An epoch-by-epoch approach was introduced to determine presence or absence of AF. Epoch length was set to 60 seconds with 50% overlap rate

(30-second sliding interval). If the sum of valid IPIs within the epoch exceeded 30 seconds, such an epoch was considered of sufficient quality to generate a classification of arrhythmia. Median AF probability associated with each valid IPI within an epoch was compared with a threshold and used to identify presence or absence of AF. When the quality of the epoch was insufficient (eg, the sum of valid IPIs <30 seconds), the epoch was classified as an "unknown rhythm."

ECG reference

ECG was acquired to generate a reference standard for presence or absence of AF. Actiwave Cardio was used in the ECV cohort to gather ECG data at 1024 Hz, and the raw waveform was visually inspected to annotate each detected heartbeat as AF, NSR, premature atrial contraction (PAC), premature ventricular contraction (PVC), or unknown/artefacts. The 12-lead Holter recorder used in the HOL cohort provided as output the temporal location of each detected beat (R-peak), which was used to determine inter-beat intervals (IBI), and the associated classification according to proprietary rhythm classification software (Veritas; Mortara, Milwaukee, WI). An experienced study technician reviewed the ECG recordings and identified AF or other arrhythmias. Rhythm type annotations were processed to generate a ground truth of AF events. The same epoch-by-epoch approach introduced to process the output of the Markov model was applied to the rhythm labels of each IBI after time synchronization with PPG-derived IPIs. AF was defined as any 60-second epoch in which >50% of the IBIs carried an AF label. Otherwise, the epoch was labeled as non-AF. The overall data processing method is schematically represented in Figure 1.

Statistical Analysis

Training and testing

Data from the ECV and HOL cohorts were split in 2 parts (a training and test set) of approximately the same size and including the same amount of AF data. Patients were randomly assigned to either the training or test set. The training data set was used to fine-tune the AF probability threshold used to classify AF from the output of the Markov model; the test data set was used to determine the algorithm accuracy on a hold-out group of patients.

Literature algorithms

Several algorithms based on features describing IBI variability and entropy have been proposed in the literature to detect AF. In our study, we selected the normalized absolute deviation (NADev) and normalized absolute difference (NADiff), which are normalized variability measures,[26] and the coefficient of the sample entropy (COSen)[27] features to process IPI sequences and compare their AF detection accuracy against the Markov model. Algorithms based on

such features were selected because they were extensively validated using large data sets of ambulatory monitoring data (>70 Holter records) as well as the MIT-BIH AF data set, and the optimal threshold to use for AF detection was disclosed in previous studies. In addition, a multiparametric algorithm was developed by averaging a surrogate probability of AF derived from normalized absolute deviation, normalized absolute difference, and coefficient of the sample entropy. The AF probability (p^{AF}_f) for each feature (f) was determined by applying a logistic function to the normalized difference between the feature value (f(i), where i indicates the current epoch) and the threshold (Th^{AF}_f) for AF detection (as disclosed in the literature; Equation 1).

$$p^{AF}_f(i) = \frac{1}{1 + e^{-kx_f(i)}} \tag{1}$$

$$x_f(i) = \frac{f(i) - Th^{AF}_f}{max_f - min_f} \tag{2}$$

Normalization was based on the feature range ($max_f - min_f$) obtained from the ECV and HOL training data set (see Equation 2), and k indicated the steepness of the logistic function curve which was set to 3.5. AF was detected when the average combined probability exceeded 0.5.

Accuracy and coverage estimation

AF classification was evaluated by comparing the algorithm output with the ground truth of rhythm annotations. Epoch-by-epoch comparison led to define accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and coverage of the classification method. Coverage was defined as the percentage of monitoring time during which a classification of either AF or non-AF could be made. Episode and duration accuracy was also determined according to the algorithm specified by the Association for the Advancement of Medical Instrumentation standards for ambulatory ECG analyzers.[28] All episodes <1 minute were excluded from episode and duration statistics. Results are reported in 2 ways: gross and average statistics. The gross method calculates the statistics by aggregating all records into one. The average method reports the mean of statistics derived for each individual record. While processing the healthy control data set, given the absence of a rhythm annotation ground truth, the records were all considered as corresponding to non-AF. Data for not-normally distributed variables are represented as median and range; otherwise, mean±SD was used.

Results

Among prospectively enrolled study participants, 6 were excluded from the HOL group and 2 from the ECV group

Table 2. Patients' Characteristics Included in the ECV and HOL Cohort

	ECV (N=18)		HOL (N=34)	
	M±SD	(Min–Max)	M±SD	(Min–Max)
Baseline characteristics				
N (%) of males	10 (56)		21 (62)	
Age, y	73.1±11.6	(45.0–87.0)	67.4±12.1	(34.0–87.0)
Weight, kg	83.3±15.3	(57.0–118.0)	85.4±22.0	(52.0–149.0)
Height, m	1.75±0.11	(1.54–1.90)	1.72±0.09	(1.51–1.86)
BMI, kg/m^2	27.2±4.1	(20.2–35.2)	28.7±6.0	(20.2–48.1)
Monitoring, h	2.4±0.3	(1.9–2.8)	25.2±4.7	(21.9–39.2)
Comorbidities, n (%)				
Hypertension	5 (28)		13 (38)	
Hyperlipidemia	2 (11)		0 (0)	
Diabetes mellitus	1 (6)		3 (9)	
Coronary artery diseases	3 (17)		5 (15)	
Heart failure	0 (0)		2 (6)	
Valve disease	0 (0)		1 (3)	
Stroke	0 (0)		3 (9)	
Medication, n (%)				
Beta-blocker	11 (61)		20 (59)	
Calcium-channel blockers	2 (11)		10 (29)	
Statin	6 (33)		15 (44)	
Antiarrhythmic	6 (33)		12 (35)	
Digoxin	2 (11)		4 (12)	
Oral anticoagulant	18 (100)		30 (88)	

(Min–Max) indicates range of values; BMI, body mass index; ECV, cohort of patients undergoing elective electrical cardioversion; HOL, cohort of patients prescribed for a Holter test; M, mean value; SD, standard deviation.

because of having rhythms other than AF, including atrial flutter or atrial tachycardia. Characteristics of study participants are presented in Table 2. ECG and wrist PPG data were collected for an average of 42 hours (21 hours of AF recordings) in the ECV and 855 hours (323 hours of AF recordings) in the HOL cohort, respectively. An example of the collected PPG data during AF and NSR is shown in Figure 2. Among participants undergoing cardioversion, 3.5% (113 beats [2–756]) of the beats (median [min–max]) were annotated as PACs and 0% (0 beats [0–1527]) as PVCs. In the HOL trial, 14 participants were in persistent AF throughout the recording period, whereas 22 had NSR. Among participants in NSR, 0.11% (94 beats [5–24 229]) and 0.02% (24 beats [0–7148]) were PACs and PVCs, respectively. Anthropometric characteristics of healthy volunteers recruited in 5 observational trials used to create the healthy control data set are presented in Table 3. In this data set, a total of around 1 year of wrist PPG data (8815 hours) was available for analysis.

AF probability threshold of the Markov model was established from the receiver operating characteristics (ROC) plot by searching in the region with >95% sensitivity which threshold conferred the largest combined specificity measured as the average value between the ECV and HOL training set. Using this threshold, AF detection accuracy was 96.3±6.8% in the ECV training data set (sensitivity=97.3±3.7%; specificity=94.3±13.6%) and 98.9±2.8% in the HOL training data set (sensitivity=97.7±4.2%; specificity=99.5±1.7%). Table 4 summarizes the accuracy obtained on the hold-out test data set for the ECV and HOL cohorts. The Markov model offered a larger ROC area under the curve as compared with conventional algorithms (NADev, NADiff, and COSen) and the multiparametric model on both the ECV and HOL test data sets (Figure 3). As a result of motion artefacts in the HOL data set, coverage for AF detection reached 56±16% (35–95%) during the home monitoring period. According to the Association for the Advancement of Medical Instrumentation standards, the accuracy of the AF detection algorithm showed a
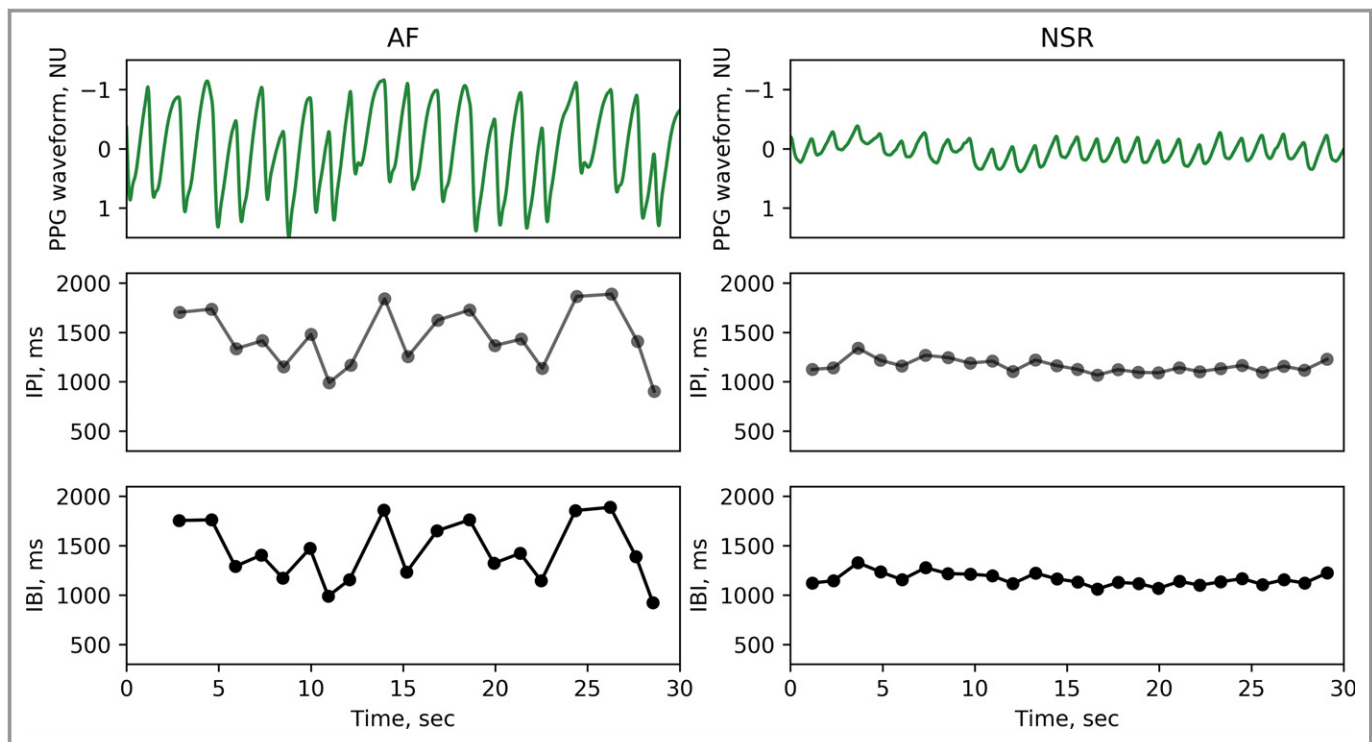
Figure 2. Representative example of photo-plethysmography (PPG) waveform, PPG-derived inter-pulse intervals (IPI), and corresponding ECG-derived inter-beat intervals (IBI) during atrial fibrillation (AF; left) and normal sinus rhythm (NSR; right) in the Holter data set (HOL).

large sensitivity and positive predictive value on both episode and duration statistics (Table 4). To validate our algorithm, we also examined the performance of the AF detection Markov model using ECG data from the MIT-BIH AF data set. The algorithm demonstrated 94% accuracy in terms of average and gross statistical performance (Table 5).

Among study participants free from arrhythmia, the Markov model algorithm very rarely detected pulse irregularities attributable to AF. Only 811 epochs were labeled as AF and those occurred in only 25 of 120 subjects (21%). Overall, this would correspond to an estimated <0.2% false positive rate in daily living conditions (6.7 hours of AF/367-day monitoring period). Other published algorithms and the multiparametric model showed a much larger number of epochs labeled as AF (from approximately 2 up to 50 times more false positives) on the same data set of healthy controls (Figure 4).

Discussion

We demonstrated that a novel algorithm analyzing pulse data from a wrist-based wearable device can accurately detect pulse irregularities attributable to AF in both hospital conditions as well as in daily life. The AF detection algorithm showed superior performance compared with other published algorithms designed to measure variability and entropy of heartbeat sequences. Furthermore, the specificity of the Markov model, and its ability to minimize false positive AF

detection (0.2%), was qualitatively verified on a large data set of apparently healthy subjects. As reported in previous studies examining pulse recordings from wrist-based wearables, motion artefacts were common and frequently corrupted PPG waveforms. Because the AF detection algorithm was designed to analyze only uncorrupted pulse recordings, AF detection coverage was, on average, approximately half (56%) of the recording time among participants wearing the wrist-based device in the home setting.

Previous studies presented the accuracy of various techniques to unobtrusively record PPG and identify occurrence of AF from pulse irregularities. Such different techniques could be categorized into solutions aimed at either instantaneous or continuous patient assessment. Applications using data provided by the camera sensor and lamp of a smartphone have been extensively validated to record pulse rate variability at the fingertip and identify AF during a spot check.[29,30] Continuous PPG measurements have been also investigated in very recent studies using earlobe-, finger-, and wrist-wearable sensors.[31–33] Accuracy of these solutions has been tested mostly under controlled hospital conditions, and AF detection was mainly derived from information carried by variability and entropy measures of IPI sequences. We recently showed that PPG measurements in hospital conditions are substantially different from those carried out in ambulatory settings. Statistical features of IPI sequences trained to classify AF in hospitalized patients did not show

Table 3. Subjects Characteristics of Volunteers Included in Healthy Control Data Set

Trial	No.	M (%)	Age, y	Weight, kg	Height, m	BMI, kg/m²	Monitoring, Hours	Description
I	18	10 (56)	32.9±8.5 (22.0–54.0)	69.4±12.0 (49.0–95.8)	1.75±0.11 (1.52–1.89)	22.6±2.0 (19.4–26.8)	68.6±8.9 (48.4–75.3)	Healthy adults monitored continuously for 3 days
II	17	9 (53)	49.7±10.1 (26.0–66.0)	81.3±10.9 (57.0–100)	1.78±0.07 (1.63–1.87)	25.7±2.8 (20.9–29.9)	13.2±4.2 (7.4–17.6)	Healthy adults monitored during sleep for 1 or 2 nights
III	24	11 (46)	39.4±14.7 (18.0–64.0)	74.8±11.5 (53.0–91.0)	1.75±0.11 (1.58–1.96)	24.6±4.5 (17.7–34.4)	23.5±1.4 (16.8–24.0)	Healthy adults monitored for 1 day
IV	46	6 (13)	38.6±10.7 (24.0–64.0)	66.9±14.2 (45.3–99.4)	1.65±0.08 (1.50–1.90)	24.7±4.7 (17.9–35.0)	78.3±21.4 (26.8–111.7)	Clinical professionals monitored for 3 days
V	15	15 (100)	49.7±7.3 (35.0–62.0)	93.1±13.5 (65.0–109)	1.81±0.07 (1.66–1.92)	28.5±4.3 (19.6–34.8)	216.3±121.3 (107.9–461.7)	Truck drivers monitored for 5 to 7 days and repeated up to 3 times
ALL	120	51 (43)	40.7±12.2 (18.0–66.0)	73.7±15.2 (45.3–109)	1.72±0.11 (1.50–1.96)	24.9±4.3 (17.7–35.0)	69.8±69.0 (7.4–461.7)	

No. number of subjects; data are expressed as mean±SD (Min–Max); ALL, statistics of the combined data sets including the 5 different observational trials (I–V). BMI indicates body mass index.

the same performance for AF detection using signals recorded outside of controlled hospital environments.[22] Similar findings were reported in a recent study evaluating the AF detection accuracy of a deep learning algorithm for a commercially available smartwatch.[34] This illustrates the importance of testing PPG-based methodologies among participants at home for properly evaluating AF detection accuracy. A state-of-the-art artificial intelligence algorithm optimized to process patient data before and after cardioversion showed only modest AF diagnostic accuracy in daily living conditions as compared with our study.[34] This suggests that analyzing pulse irregularity in the absence of motion represents a more-reliable way to detect AF than processing sequences of average heart rate from a smartwatch.

Our AF detection algorithm consistently and accurately discriminated AF from normal rhythm in both controlled hospital as well as home conditions. The Markov model accuracy was further tested with the MIT-BIH AF data set, and, compared with other published algorithms, it demonstrated somewhat lower sensitivity, but among the best accuracy and positive predictive value. AF detection accuracy of conventional variability and entropy features was consistent with that reported previously.[26,27] However, these literature algorithms, when applied to the healthy control data set, showed a detrimentally large number of false positives as compared to the Markov model. Fine-tuning their decision threshold for AF detection using PPG data, and perhaps combining and expanding the number of features in a single detection algorithm,[35] could improve the specificity.

Clinical Relevance

The gold standard for AF diagnosis remains the ECG.[9,12] Yet, noninvasive ECG monitoring devices available today are cumbersome, costly, and suffer from poor patient adherence. These characteristics limit our capacity to perform long-term noninvasive investigation for incident or recurrent paroxysmal AF.[36,37] A convenient wrist-wearable PPG sensor able to detect pulse irregularities attributable to AF could represent a valid solution to assist with screening for AF by providing a comfortable method to identify individuals who might benefit from confirmatory diagnostic ECG testing.

Opportunistic screening of older adults carrying significant risk factors for stroke, such as hypertension, diabetes mellitus, and sleep apnea, may benefit from long-term and unobtrusive rhythm monitoring by PPG to support stroke prevention programs. The cost-effectiveness of such a solution would require a proper clinical evaluation. The first evidence for pulse-based AF screening in stroke prevention programs has been provided for an oscillometric blood pressure device[38] able to automatically determine pulse irregularity that may be caused by AF.[39]

Table 4. Accuracy and Coverage of the Algorithm Designed to Detect Pulse Irregularities Attributable to AF According to the ECV and HOL Testing Data Set

Testing Data Set	Epoch-by-Epoch Statistics					Episode Statistics			Duration Statistics	
	Accuracy % (CI)	Sensitivity % (CI)	Specificity % (CI)	NPV % (CI)	PPV % (CI)	Coverage % (CI)	Sensitivity % (CI)	PPV % (CI)	Sensitivity % (CI)	PPV % (CI)
ECV										
Average	98±4 (96–100)	97±7 (91–100)	100±1 (99–100)	98±6 (93–100)	99±4 (96–100)	57±18 (46–68)	98±6 (94–100)	99±4 (96–100)	98±7 (93–100)	94±13 (86–99)
Gross	98	96	100	98	100	47	99	99	96	98
HOL										
Average	97±10 (92–100)	93±14 (82–100)	100±0 (100–100)	n.a.	n.a.	52±12 (47–57)	95±11 (85–99)	n.a.	93±14 (81–99)	n.a.
Gross	97	93	100	95	100	48	95	100	93	100

Sensitivity and PPV for episode and duration statistics are defined according to the Association for the Advancement of Medical Instrumentation standards. AF indicates atrial fibrillation; CI, 95% confidence interval of the mean as obtained by bootstrapping and resampling of the test data set (100 iterations); ECV, cardioversion cohort of patients; Average, indicates accuracy described by mean (± SD deviation) of the value across patients; Gross, accuracy determined by aggregating results from each patient into one; HOL, Holter cohort of patients; NPV, negative predictive value; PPV, positive predictive value; n.a., not available. NPV and PPV could not be defined as *average* accuracy in epoch-by-epoch statistics because patients showed either AF or non-AF in the reference annotations in the HOL cohort.

Ambulatory rhythm monitoring with PPG may also be used to better classify a patient's condition into paroxysmal, persistent, or permanent AF. Clinical distinction between AF categories is needed to personalize treatment and select appropriate therapies for patients.[9] However, data from implantable devices show a large discrepancy between clinical classification of AF patients and the actual AF burden or temporal persistence.[37] Charitos et al[37] observed that in a 1-year follow-up, only 47% of paroxysmal patients actually suffered from short and self-terminating episodes of AF. Likewise, barely 33% of patients clinically classified as persistent AF showed arrhythmic events >7 consecutive days. This suggests that a wrist-wearable PPG sensor could remove the discordance between clinical assessment of AF and actual AF persistence, and perhaps lead to better identification of patients who would likely respond to the appropriate therapy.

Objective assessment of AF burden has also been stressed as important to confer the correct stroke risk to AF patients. Indeed, a clear trend of increasing stroke risk from paroxysmal to persistent to permanent AF has been suggested from the ACTIVE-A (Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events) and AVERROES (Apixaban Versus Acetylsalicylic Acid [ASA] to Prevent Stroke in Atrial Fibrillation Patients Who Have Failed or Are Unsuitable for Vitamin K Antagonist Treatment) trials.[40] Yet, clinical guidelines for stroke prevention do not take the AF burden into account to gauge stroke risk in patients, in part because of the lack of data from long-term rhythm monitors.[41] In the near future, convenient and broadly applicable solutions for measuring AF burden using PPG may pave the way to generate fresh insights into relative stroke risk among patients with brief episodes of AF compared with individuals with longer episodes or greater burden of AF.

Clinical relevance of transient, paroxysmal AF detected during acute conditions, such as infection, pneumonia, acute myocardial infarction, or surgery (so-called secondary AF), is unknown. However, recent studies[42] suggest that secondary AF may have clinical significance[43] despite the commonly held notion that secondary AF is caused by a reversible precipitant.[41] Evidence indicates that around half of patients suffering from this type of transient AF would manifest episodes of arrhythmia in a 5-year period.[42] In this patient population, a sensitive and convenient method to detect pulse irregularities attributable to AF over a long-term period may help facilitate risk stratification for stroke and health deterioration.

Limitations

A limitation of this study was the lack of transient and short AF episodes in daily living conditions. The HOL cohort only included 14 participants with persistent AF (100% of recording windows). Thus, we could not determine the
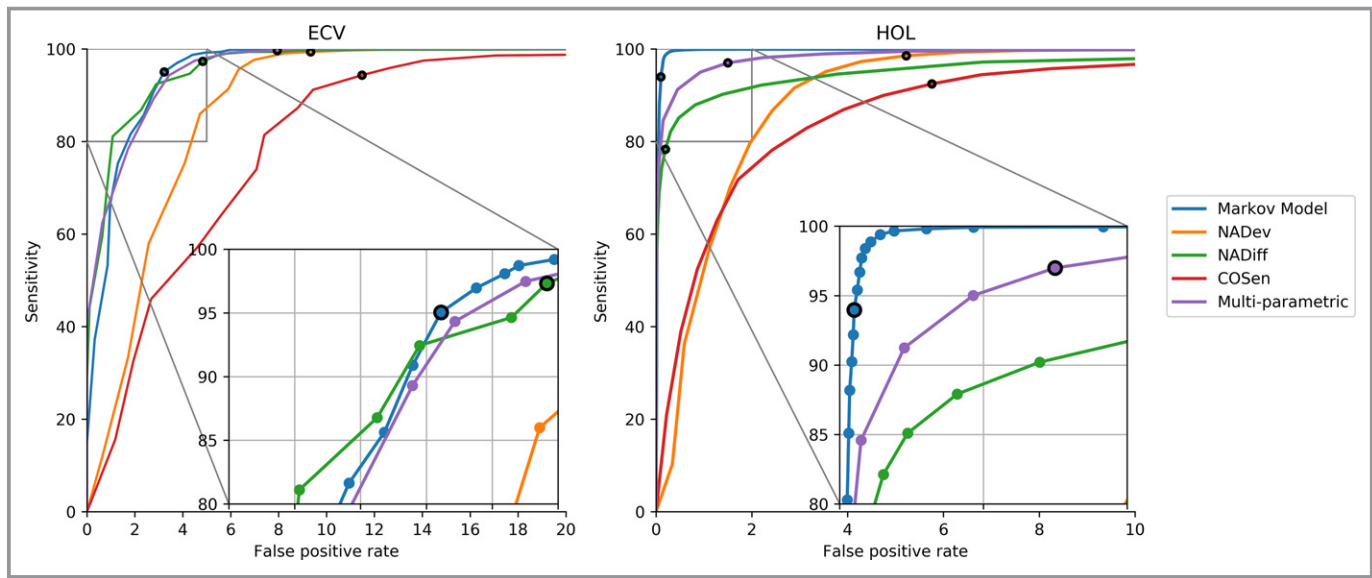
Figure 3. ROC curve for the atrial fibrillation detection algorithm (Markov model), selected literature algorithms (NADev, NADiff, and COSen) and a multiparametric algorithm (combining NADev, NADiff, and COSen) for the cardioversion (ECV) and Holter (HOL) cohort. Black circles indicate the operative point of the different methods. COSen indicates coefficients of the sample entropy; ECV, electrical cardioversion; NADev, normalized absolute deviation; NADiff, normalized absolute difference.

impact of loss of monitoring coverage observed during the day on detection accuracy of intermittent AF. Yet, the sensitivity of the presented Markov model for detection of transient AF was validated using the MIT-BIH AF data set, albeit from ECG data. Future studies are needed to

objectively evaluate the clinical implications of low monitoring coverage for AF detection despite the long-term monitoring modality offered by the technology. A second limitation was the absence of a rhythm annotation reference for the PPG recordings used to test the false positive rate of

Table 5. Accuracy of the Markov Model Designed to Detect Pulse Irregularities Attributable to AF and of Literature Algorithms According to MIT-BIH AF Data Set

Algorithm/Feature	Accuracy % (CI)	Sensitivity % (CI)	Specificity % (CI)	PPV % (CI)	NPV % (CI)
Markov model					
Average	94±12 (89–98)	84±21 (77–90)	98±5 (96–100)	90±21 (80–96)	92±14 (87–97)
Gross	94	86	99	98	91
NADev					
Average	94±6 (92–96)	95±6 (93–98)	89±17 (82–95)	67±36 (52–80)	97±6 (95–99)
Gross	94	98	92	89	99
NADiff					
Average	92±13 (86–97)	92±16 (85–97)	92±15 (87–97)	80±30 (69–93)	95±11 (91–99)
Gross	92	91	93	89	94
COSen					
Average	91±9 (85–94)	89±10 (85–92)	88±16 (82–94)	66±35 (48–78)	91±16 (86–97)
Gross	91	91	90	86	94
Multiparametric					
Average	95±8 (91–97)	96±7 (94–99)	91±15 (85–97)	73±31 (57–84)	98±4 (96–100)
Gross	95	98	93	90	98

Average, indicates accuracy described by mean ± SD of the value across patients; Gross, accuracy determined by aggregating results from each patient into one; AF indicates atrial fibrillation; CI, 95% confidence interval of the mean as obtained by bootstrapping and resampling of the data set (100 iterations); NPV, negative predictive value; PPV, positive predictive value. Average indicates accuracy described by meanSD of the value across patients. Any AF event <1 minute was removed from the analysis, and 1 record (5091) was discarded because AF was only present for <1 minute.
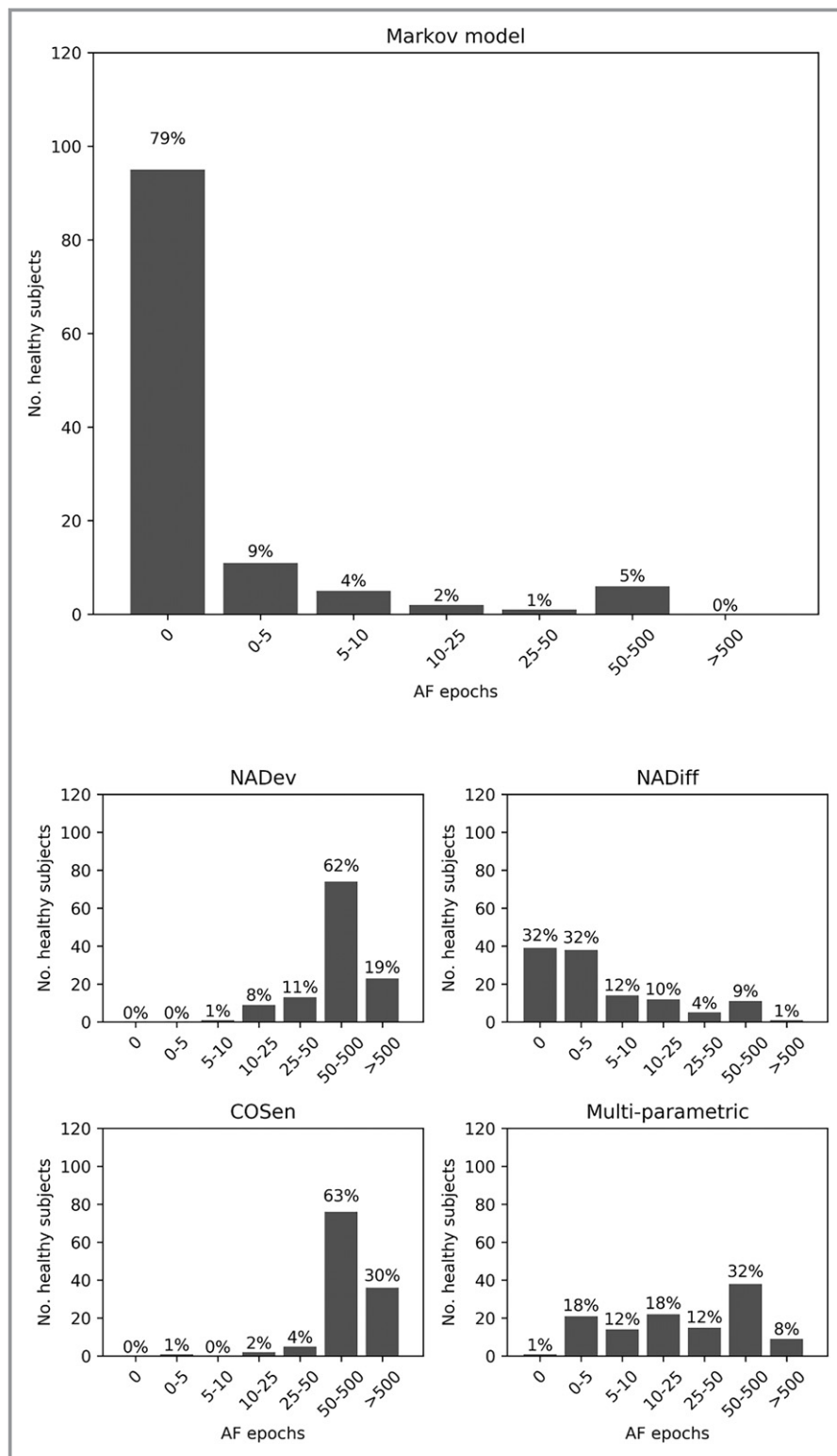
Figure 4. Histogram indicating the number of subjects for each categorical amount of atrial fibrillation (AF) epochs as output by the AF detection algorithm (Markov model), literature algorithms (NADev, NADiff, and COSen) and a multiparametric algorithm (combining NADev, NADiff, and COSen) with the control data set of apparently healthy subjects. COSen indicates coefficients of the sample entropy; NADev, normalized absolute deviation; NADiff, normalized absolute difference.

the AF detection algorithm in the healthy control data set. In this study, we assumed that volunteers recruited in the different observational studies were not suffering from AF. It is possible that these participants experienced brief episodes of undetected AF given that they were not wearing contemporaneous Holter monitors. Another important limitation of this study was that the diagnostic accuracy of the Markov model for AF discrimination is unproven in populations affected by other cardiac rhythm disorders, such as atrial flutter or tachycardia. Among the few dropouts of the HOL cohort suffering from continuous atrial flutter, we noticed that AF was rarely detected in some patients (2 patients, specificity: 96.4% and 92.1% with coverage >50%), but often in others (3 patients, specificity: 7.5%, 17.6%, and 2.8% with coverage >50%). In addition, the Markov model detected AF only occasionally in the dropout patient from the HOL trial suffering from continuous atrial tachycardia (specificity, 78.8% with coverage >50%). Follow-up studies are needed to collect more data from large cohorts of free-living individuals affected by paroxysmal AF and other cardiac rhythm abnormalities to confirm the algorithm's performance for AF detection.

Conclusion

A novel algorithm analyzing pulse data from a wrist-wearable device can accurately detect pulse irregularities associated with AF. Our findings suggest that wrist-based wearables capable of pulse monitoring can be used as long-term monitors for AF and thereby enhance detection of paroxysmal AF and assessment of AF burden among patients affected by this dysrhythmia.

Acknowledgments

Sources of Funding

Disclosures

Bonomi, Margarito, van Dinther, Muesch, de Morree, and Aarts are employed by Philips Research. Schipper and Babaeizadeh are employed by Philips Healthcare.

References

1. Friberg L, Bergfeldt L. Atrial fibrillation prevalence revisited. *J Intern Med*. 2013;274:461–468.

2. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim YH, McAnulty JH, Zheng ZJ, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M, Murray CJ. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*. 2014;129:837–847.

3. Thrall G, Lane D, Carroll D, Lip GY. Quality of life in patients with atrial fibrillation: a systematic review. *Am J Med*. 2006;119:448.e1–448.e19.

4. Benjamin EJ, Wolf PA, Agostino RB, Silbershatz H, Kannel WB, Levy D. Clinical investigation and reports impact of atrial fibrillation on the risk of death the Framingham Heart Study. *Circulation*. 1998;98:946–952.

5. Wolf PA, Dawber TR, Thomas HE, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham study. *Neurology*. 1978;28:973–977.

6. Wang TJ, Larson MG, Levy D, Vasan RS, Leip EP, Wolf PA, D'Agostino RB, Murabito JM, Kannel WB, Benjamin EJ. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation*. 2003;107:2920–2925.

7. Steinberg BA, Kim S, Fonarow GC, Thomas L, Ansell J, Kowey PR, Mahaffey KW, Gersh BJ, Hylek E, Naccarelli G, Go AS, Reiffel J, Chang P, Peterson ED, Piccini JP. Drivers of hospitalization for patients with atrial fibrillation: results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). *Am Heart J*. 2014;167:735–742.

8. Amin AN, Jhaveri M, Lin J. Hospital readmissions in US atrial fibrillation patients. *Am J Ther*. 2013;20:143–150.

9. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener HC, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, Popescu BA, Schotten U, Van Putte B, Vardas P, Agewall S, Camm J, Esquivias GB, Budts W, Carerj S, Casselman F, Coca A, De Caterina R, Deftereos S, Dobrev D, Ferro JM, Filippatos G, Fitzsimons D, Gorenek B, Guenoun M, Hohnloser SH, Kolh P, Lip GYH, Manolis A, McMurray J, Ponikowski P, Rosenhek R, Ruschitzka F, Savelieva I, Sharma S, Suwalski P, Tamargo JL, Taylor CJ, Van Gelder IC, Voors AA, Windecker S, Zamorano JL, Zeppenfeld K. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J*. 2016;37:2893–2962.

10. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med*. 2007;146:857–867.

11. Brachmann J, Morillo CA, Sanna T, Di Lazzaro V, Diener HC, Bernstein RA, Rymer M, Ziegler PD, Liu S, Passman RS. Uncovering atrial fibrillation beyond short-term monitoring in cryptogenic stroke patients: three-year results from the cryptogenic stroke and underlying atrial fibrillation trial. *Circ Arrhythm Electrophysiol*. 2016;9:1–10.

12. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, Conti JB, Ellinor PT, Ezekowitz MD, Field ME, Murray KT, Sacco RL, Stevenson WG, Tchou PJ, Tracy CM, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Creager MA, Curtis LH, DeMets D, Guyton RA, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen WK, Yancy CW. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *Circulation*. 2014;130:e199–e267.

13. Jahangir A, Lee V, Friedman PA, Trusty JM, Hodge DO, Kopecky SL, Packer DL, Hammill SC, Shen WK, Gersh BJ. Long-term progression and outcomes with aging in patients with lone atrial fibrillation: a 30-year follow-up study. *Circulation*. 2007;115:3050–3056.

14. Rosero SZ, Kutyifa V, Olshansky B, Zareba W. Ambulatory ECG monitoring in atrial fibrillation management. *Prog Cardiovasc Dis*. 2013;56:143–152.

15. Camm AJ, Corbucci G, Padeletti L. Usefulness of continuous electrocardiographic monitoring for atrial fibrillation. *Am J Cardiol*. 2012;110:270–276.

16. Ziegler PD, Koehler JL, Mehra R. Comparison of continuous versus intermittent monitoring of atrial arrhythmias. *Heart Rhythm*. 2018;3:1445–1452.

17. Benjamin EJ, Chen PS, Bild DE, Mascette AM, Albert CM, Alonso A, Calkins H, Connolly SJ, Curtis AB, Darbar D, Ellinor PT, Go AS, Goldschlager NF, Heckbert SR, Jalife J, Kerr CR, Levy D, Lloyd-Jones DM, Massie BM, Nattel S, Olgin JE, Packer DL, Po SS, Tsang TSM, Van Wagoner DR, Waldo AL, Wyse DG. Prevention of atrial fibrillation: report from a National Heart, Lung, and Blood Institute Workshop. *Circulation*. 2009;119:606–618.

18. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*. 2000;101:e215–e220.

19. Valenti G, Westerterp KR. Optical heart rate monitoring module validation study. In: IEEE International Conference on Consumer Electronics. New York, NY: IEEE. 2013:195–196.

20. Bonomi AG, Goldenberg S, Papini G, Kraal J, Stut W, Sartor F, Kemps H. Predicting energy expenditure from photo-plethysmographic measurements of heart rate under beta blocker therapy: data driven personalization strategies based on mixed models. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:7642–7646.

21. Van Andel J, Ungureanu C, Aarts R, Leijten F, Arends J. Using photoplethysmography in heart rate monitoring of patients with epilepsy. *Epilepsy Behav*. 2015;45:142–145.

22. Eerikäinen LM, Dekker L, Bonomi AG, Vullings R, Schipper F, Margarito J, De Morree HM, Aarts RM. Validating features for atrial fibrillation detection from photoplethysmogram under hospital and free-living conditions. In 44th Computing in Cardiology Conference, CinC 2017 (pp. 3–6), September 24–27, 2017 Rennes, France.

23. Bonomi AG, Schipper F, Eerikäinen LM, Margarito J, Aarts RM, Babaeizadeh S, de Morree HM, Dekker L. Atrial fibrillation detection using photo-plethysmography and acceleration data at the wrist. In 43rd Computing in Cardiology Conference, CinC 2016 (pp. 277–280), September 11–14, 2016, Vancouver, British Columbia, Canada .

24. Moody GB, Mark RG. A new method for detecting atrial fibrillation using RR intervals. In 10th Annual Computing in Cardiology Conference (pp. 227–230), October 4–7, 1983, Aachen, Germany. Computers in Cardiology. 1983:227–230.

25. Babaeizadeh S, Gregg RE, Helfenbein ED, Lindauer JM, Zhou SH. Improvements in atrial fibrillation detection for real-time monitoring. *J Electrocardiol*. 2009;42:522–526.

26. Ghodrati A, Murray B, Marinello S. RR interval analysis for detection of atrial fibrillation in ECG monitors. *Conf Proc IEEE Eng Med Biol Soc*. 2008;2008:601–604.

27. Lake DE, Moorman JR. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am J Physiol Heart Circ Physiol*. 2011;300:H319–H325.

28. American National Standards Institute. ANSI/AAMI EC38:2007. Medical electrical equipment. Arlington, VA: American National Standards Institute; 2007.

29. McManus DD, Lee J, Maitas O, Esa N, Pidikiti R, Carlucci A, Harrington J, Mick E, Chon KH. 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.

30. Chan PH, Wong CK, Poh YC, Pun L, Leung WW, Wong YF, Wong MM, Poh MZ, Chu DW, Siu CW. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. *J Am Heart Assoc*. 2016;5:e003428. DOI: 10.1161/JAHA.116.003428.

31. Corino VDA, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. *Physiol Meas*. 2017;38:787–799.

32. Conroy T, Guzman JH, Hall B, Tsouri G, Couderc JP. Detection of atrial fibrillation using an earlobe photoplethysmographic sensor. *Physiol Meas*. 2017;38:1906–1918.

33. Tang SC, Huang PW, Hung CS, Shan SM, Lin YH, Shieh JS, Lai DM, Wu AY, Jeng JS. Identification of atrial fibrillation by quantitative analyses of fingertip photoplethysmogram. *Sci Rep*. 2017;7:45644.

34. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, Vittinghoff E, Lee ES, Fan SM, Gladstone RA, Mikell C, Sohoni N, Hsieh J, Marcus GM. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol*. 2018;3:409–416.

35. Dash S, Chon KH, Lu S, Raeder EA. Automatic real time detection of atrial fibrillation. *Ann Biomed Eng*. 2009;37:1701–1709.

36. Kishore A, Vail A, Majid A, Dawson J, Lees KR, Tyrrell PJ, Smith CJ. Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. *Stroke*. 2014;45:520–526.

37. Charitos EI, Pürerfellner H, Glotzer TV, Ziegler PD. Clinical classifications of atrial fibrillation poorly reflect its temporal persistence: insights from 1,195 patients continuously monitored with implantable devices. *J Am Coll Cardiol*. 2014;63:2840–2848.

38. Willits I, Keltie K, Craig J, Sims A. WatchBP Home A for opportunistically detecting atrial fibrillation during diagnosis and monitoring of hypertension: a NICE Medical Technology Guidance. *Appl Health Econ Health Policy*. 2014;12:255–265.

39. Stergiou GS, Karpettas N, Protogerou A, Nasothimiou EG, Kyriakidis M. Diagnostic accuracy of a home blood pressure monitor to detect atrial fibrillation. *J Hum Hypertens*. 2009;23:654–658.

40. Vanassche T, Lauw MN, Eikelboom JW, Healey JS, Hart RG, Alings M, Avezum A, Diaz R, Hohnloser SH, Lewis BS, Shestakovska O, Wang J, Connolly SJ. Risk of ischaemic stroke according to pattern of atrial fibrillation: analysis of 6563 aspirin-treated patients in ACTIVE-A and AVERROES. *Eur Heart J*. 2015;36:281–288.

41. Mcintyre W, Healey J. Stroke prevention for patients with atrial fibrillation: beyond the guidelines. *J Atr Fibrillation*. 2017;9:1–7.

42. Lubitz SA, Yin X, Rienstra M, Schnabel RB, Walkey AJ, Magnani JW, Rahman F, McManus DD, Tadros TM, Levy D, Vasan RS, Larson MG, Ellinor PT, Benjamin EJ. Long-term outcomes of secondary atrial fibrillation in the community: the Framingham Heart Study. *Circulation*. 2015;131:1648–1655.

43. Walkey AJ, Hogarth DK, Lip GY. Optimizing atrial fibrillation management: from ICU and beyond. *Chest*. 2015;148:859–864.

SUPPLEMENTAL MATERIAL

Data S1.

MARKOV MODEL TO DETECT ATRIAL FIBRILLATION FROM PULSE INTERALS

This section describes the details of the Markov model used to determine the likelihood of AF from a sequence of inter-pulse-intervals (IPI). The IPI sequence is modeled as an 11-states first-order Markov process in which each observed IPI is associated to one of 11 different possible states. This association is done firstly by determining the ratio between the observed IPI and the average IPI in the sequence, and secondly by classifying such ratio into 11 categories of values ranging from 0.75 to 1.25 at steps of 0.05. This corresponds to identify to which extent the observed IPI is shorter, similar, or larger than average. In accordance with the work of Moody and Mark[1], the average IPI ($\overline{IPI}$) was determined recursively as follows:

$$\overline{IPI}(i) = 0.75 \times \overline{IPI}(i-1) + 0.25 \times IPI(i)$$

ECG recordings were used in a previous study[2] as training data to define the statistics of transition from one state to another and derive transition probability matrixes in either presence or in absence of AF. Assuming a sequence $\mathbf{X}$ of n states:

$$\mathbf{X} = \{x_1, x_2, x_3, \dots x_n\} \quad (x_i \, in \, \{s_{75}, s_{80}, s_{85}, \dots s_{125}\})$$

where s indicates one of the 11-states of the model, the transition probability matrix in presence of AF or absence of AF (*not_AF*) can be defined as:

$$p_{j,k,AF} = P\big(x_j \mid x_k, AF\big)$$

$$p_{j,k,not\_AF} = P(x_j \mid x_k, not\_AF)$$

Where x_j and x_k are two consecutive states in the process and $p_{j,k,R}$ their relative frequency during a certain heart rhythm R (*AF* or *not_AF*). From this we can derive the conditional probability of **X** given the process R (either *AF* or *not_AF*):

$$P(X \mid R) = \prod_{i=1}^{n-1} p_{i+1,i,R}$$

The rhythm R that maximizes P indicates the maximum likelihood procedure for the observed sequence of states **X**. As suggested by Moody and Mark[1], the logarithm of the transition matrix for not_AF was divided by that of AF and used to determine the Markov score (S_i) iteratively by adding each element in the resulting matrix according to the transitions between the n observed states **X**.

$$S_i = \sum_{i=1}^{n-1} \log\left(\frac{p_{i+1,i,not\_AF}}{p_{i+1,i,AF}}\right)$$

In our work n was kept equal to the entire length of the IPI sequence. As a final step the Markov score was processed using a previously published averaging filter[1]. The resulting Markov score (S_i) tended to decrease below zero in presence of AF and such value was compared to a threshold to make a binary decision between presence and absence of AF. The Markov model which was designed and trained using ECG data was optimized in our study to process IPI sequences from PPG for AF detection. Such optimization focused solely on fine-tuning the probability threshold applied to the Markov score (S_i) to identify episodes of AF.

Supplemental References:

1. Moody GB, Mark RG. A New Method for Detecting Atrial Fibrillation Using RR Intervals. *Computers in Cardiology*. 1983. p. 227–230.

2. Babaeizadeh S, Gregg RE, Helfenbein ED, Lindauer JM, Zhou SH. Improvements in atrial fibrillation detection for real-time monitoring. *J Electrocardiol*. 2009;42:522–526.