Using Apple Watch for Arrhythmia Detection

December 2018

Contents

Overview **3**

Introduction....3

PPG-based arrhythmia detection **3**

Technical and Feature Description....3

Preclinical Development....4

Clinical Validation....5

Apple Heart Study....5

AHS Sub-Study Experiment Design....5

Results....5

ECG-based detection **6**

Technical and Feature Description....6

Preclinical Development....6

Clinical Validation....7

Experiment Design....7

Results....8

Conclusions **11**

Overview

With the availability of watchOS 5.1.2 and Apple Watch Series 4, Apple Watch customers now have access to two features to detect heart arrhythmias such as atrial fibrillation. Apple Watch Series 1 and later can look for arrhythmias using a photoplethysmograph-based algorithm, and the ECG app on Apple Watch Series 4 is capable of generating an ECG similar to a Lead I electrocardiogram. This app also classifies an ECG as sinus rhythm (SR), atrial fibrillation (AF), or inconclusive, and reports high or low heart rate. This paper is intended to provide a more detailed understanding of the capabilities of these features, including testing and validation.

Introduction

Atrial fibrillation, a type of irregular heart rhythm in which the atria of the heart beat irregularly and sometimes rapidly, is a leading cause of stroke. However, AF is often asymptomatic, leading many individuals with AF to be unaware they have this condition. The combination of stroke risk, asymptomatic presentation, effective pharmacologic treatments minimizing stroke risk, and increasing market penetration of consumer devices with the potential to detect AF have increased interest in the early identification of AF outside the clinical setting.

With watchOS 5.1.2, Apple Watch Series 1 and later are capable of identifying periods of irregular pulse suggestive of AF using photoplethysmograph (PPG) signals combined with an algorithm. In addition to this PPG-based identification algorithm, Apple Watch Series 4 has an electrical heart sensor that, when using the ECG app, enables the generation and analysis of an ECG similar to a Lead I ECG.

PPG-based arrhythmia detection

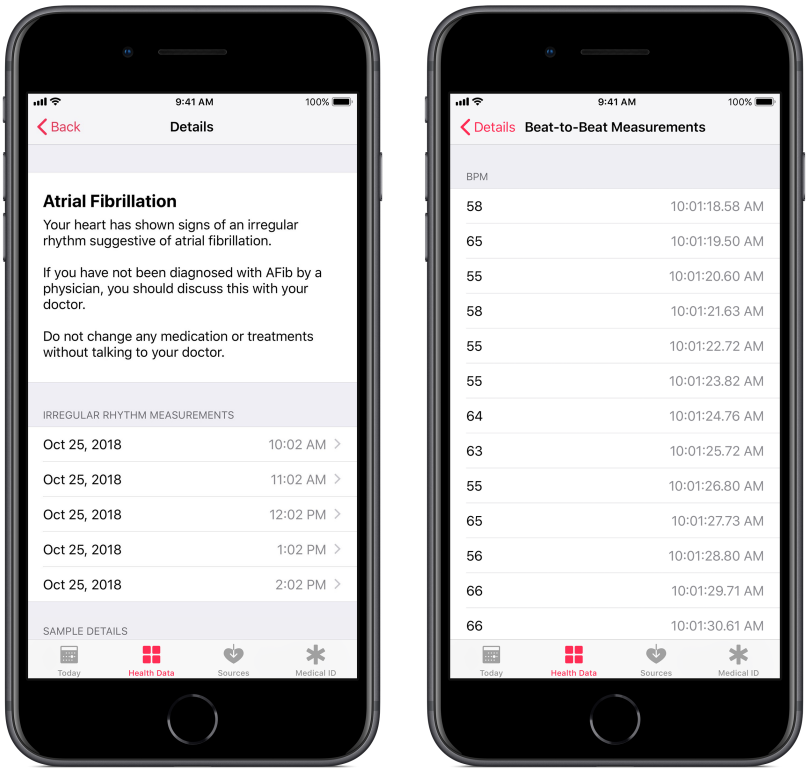
Technical and Feature Description

Apple Watch has an optical heart sensor that uses green LED lights paired with light-sensitive photodiodes to detect blood volume pulses in a user's wrist using photoplethysmography. These sensors and underlying algorithms are the basis for the heart rate and heart rate variability (HRV) detection enabled on Apple Watch Series 1 and later. To determine HRV, Apple Watch captures a tachogram, a plot of the time between heartbeats, every two to four hours. Beginning with watchOS 5.1.2, a user may also choose to enable an arrhythmia detection feature that utilizes these tachograms. To use the Irregular Rhythm Notification feature on Apple Watch, a user must first complete onboarding within the Health app on the user's paired iPhone to learn how to use the feature and receive education regarding AF. To learn more about the user experience, visit https://support.apple.com/kb/HT208931.

If the PPG-based arrhythmia detection is enabled, each tachogram is classified using a proprietary algorithm to determine if an irregular rhythm may be present. An irregular tachogram initiates a cascade of more frequent tachogram collection (as frequently as possible, subject to a minimum spacing of 15 minutes) and analysis. Tachograms are collected and analyzed only if the user remains still enough to obtain a reading; because of this, the algorithm is not always monitoring the user, but rather is doing so opportunistically when adequate signal is available for collection/analysis. If five out of six sequential

tachograms (including the initial one) are classified as irregular within a 48-hour period, the user is notified of the potential arrhythmia. In addition to the notification, the user can access more information related to these irregular tachograms within the Health app (Figure 1). If two tachograms are classified as not irregular before the threshold is reached, the cycle is reset and tachogram collection returns to the baseline rate (every two hours).

Figure 1: Health App View of Irregular Rhythm Measurements



Within the Health app, users can see the times when the algorithm identified an irregular tachogram that contributed to a notification (left). Selecting any of these specific dates/times allows a user to visualize the beat-to-beat measurements calculated from each tachogram.

Preclinical Development

Prior to clinical testing, studies were conducted to develop the PPG-based detection algorithm and to evaluate algorithm performance across a variety of conditions and user behaviors. Among these were deep breathing, riding in a car, hand tremors and motion, reduced hand/wrist perfusion, overnight wear, rapid ventricular response in individuals with AF, and other arrhythmias. These studies were performed in 2300 control subjects and more than 500 subjects with AF.

Because PPG relies on light absorptivity, the arrhythmia detection algorithm was tested across a variety of skin types and tones to ensure that sensor platform adjustments for skin tone were sufficient in the context of the algorithms used to detect arrhythmias. Melanin has high absorptivity at the wavelength used by the green LED on the Apple Watch, making PPG heart rate measurement potentially more difficult in darker skin tones. To account for this, the Apple Watch sensing platform adjusts LED current (and hence light output), photodiode gain (sensitivity to light), and sampling rate to ensure adequate signal amplitude across the full range of human skin tone.

For validation purposes, 1.3 million tachograms from 1124 subjects (51% female) with varying skin type and tone (Fitzpatrick skin type and spectrophotometer-measured skin lightness at the wrist) were analyzed. As the primary engineering concerns focused on signal amplitudes in individuals with dark skin, nearly 5% of enrolled subjects had Fitzpatrick type VI skin (data not shown), about twice the expected prevalence in the U.S. population. Validation efforts demonstrated no significant difference in algorithm sensitivity or specificity across skin types/tones.

Clinical Validation

Apple Heart Study

The Apple Heart Study (AHS) is a prospective, single-arm pragmatic study conducted virtually to evaluate the ability of the Apple Watch–based irregular pulse notification algorithm to identify arrhythmia suggestive of AF. Within AHS, if a user met the 5/6 irregular tachogram threshold, the user received an iPhone and Apple Watch notification and had the option of contacting a telehealth study physician and being sent an ambulatory ECG patch (ePatch, BioTelemetry, Inc., Conshohocken, PA). Participants were instructed to wear the ePatch for up to seven days; however, data collected from a participant were considered adequate with a minimum analyzable time of one hour.

Enrollment for AHS has concluded, and the full study results will be published elsewhere once available.

AHS Sub-Study Experiment Design

A sub-study of data collected in AHS was conducted to determine if the tachogram classification algorithm (individual or spot tachogram) and the confirmation cycle algorithm (alert-level, five out of six tachograms) have acceptable positive predictive value (PPV) as compared with the ePatch monitoring in identifying irregular rhythms consistent with AF in a subset of AHS participants. AHS investigators were aware of the sub-study, subsequent analyses, and data submission to the FDA; however, AHS investigators were blinded to the sub-study results while AHS was ongoing. The Institutional Review Board (IRB) that approved AHS determined that this sub-study was exempt from IRB oversight. All AHS participants provided informed consent, which included the use of their study data for the purposes of the sub-study.

Sub-study data were collected from AHS participants enrolled between November 30, 2017, and June 22, 2018. The subjects in this sub-study received an irregular rhythm notification from the AHS app and subsequently received and wore an ambulatory ECG patch (ePatch) for interpretation of the ambulatory ECG findings. The initial irregular tachograms leading to the first notification and potentially launching the first telehealth encounter were not analyzed as part of this sub-study; only irregular tachograms and notifications occurring while a user was wearing the study-provided ePatch were analyzed.

Two independent ECG adjudicators with U.S. board certification in cardiology and/or electrophysiology provided review and adjudication of ECG strips, classifying them as SR, AF, Other Irregular Rhythm, or Unreadable. If the adjudicators did not agree, a third, similarly qualified adjudicator evaluated the strip. These adjudicators were blinded to the tachogram classification. The adjudicator ECG classification and algorithm-determined tachogram classification were securely sent to the study statistician for data analysis.

Results

Of the 226 sub-study participants who received an initial arrhythmia notification and wore an ePatch for approximately one week, 41.6% (94/226) had AF detected by the ePatch. During concurrent wear of Apple Watch and an ePatch, 57/226 participants received an AF notification (i.e., had five out of six consecutive tachograms classified as abnormal). Of those, 78.9% (45/57) showed concordant AF on

the ePatch data and 98.2% (56/57) showed AF or other clinically relevant arrhythmias. These results demonstrate that, while in the majority of cases the notification will accurately represent the presence of AF, in some instances a notification may indicate the presence of an arrhythmia other than AF. No serious device adverse effects were observed.

ECG-based detection

Technical and Feature Description

Apple Watch Series 4 incorporates a titanium electrode in the Digital Crown and an ultrathin chromium silicon carbon nitride layer applied to the sapphire crystal on the back of Apple Watch. The ECG app reads and records the electrical impulses that control the heart from the user's fingertip (Digital Crown) and the wrist (back of Apple Watch), which creates a closed circuit. To use the ECG app on Apple Watch, a user must first complete onboarding within the Health app on the user's paired iPhone to learn how to use the feature and receive education regarding AF. To generate an ECG, a user must open the ECG app installed on Apple Watch, then apply a finger from the hand contralateral to the wrist where Apple Watch is worn to the Digital Crown for 30 seconds. Lead polarity is determined by the wrist placement of Apple Watch selected in Settings.

After obtaining the ECG, a proprietary algorithm is used to classify the ECG tracing as SR, AF, or inconclusive. This rhythm classification, average heart rate, user-reported symptoms, and waveform are all stored in HealthKit and can be shared by the user as a PDF from the Health app on the user's paired iPhone. To learn more about the user experience, visit https://support.apple.com/kb/HT208955.

Preclinical Development

The ECG signal detection and classification algorithm were also tested in multiple studies prior to commencing clinical validation. The sensors and classification algorithm were tested across various ethnicities, wrist circumferences, BMI ranges, ages, non-AF arrhythmias, degrees of band tightness, postures, and exercise states/sweating. Approximately 2000 subjects were involved in these tests, about 15% of whom had previously been diagnosed with AF or other irregular heart rhythms.

Increased frequency of "unreadable" ECG was the primary variation in algorithm performance. The factors leading to this variation were low signal amplitude (as a result of right-axis deviation, particularly noted in those with low BMI, or sweating noted during testing after exercise sessions) and motion artifacts as a result of user behavior (including shivering artifacts during testing in Fairbanks, Alaska). A higher proportion of samples were deemed unclassified after running, as elevated heart rates pushed the signals out of the range considered as normal SR (50 to 100 beats per minute). Apple Watch uses dry electrodes designed to be mechanically strong and corrosion resistant as appropriate for a wearable device; however, dry electrodes, particularly those placed on extremities, are inherently more prone to introducing noise such as that described above relative to the temporary gel electrodes used in clinical devices.

In addition to the factors mentioned above, the presence of certain non-AF arrhythmias also resulted in significantly different algorithm performance compared with subjects in sinus rhythm. These conditions and the results are described in Table 1.

Table 1. Non-AF Arrhythmias Affecting Algorithm Performance

Arrhythmia	Variation
L/R bundle branch block	1.3% of trials classified as AF
Second-degree AV block	25% of trials unclassified; of classified trials, 88.9% classified as AF
Bigeminy	All trials unclassified
Frequent ectopic beats	22.7% unclassified; of those classified, none labeled AF
Junctional rhythm	18.2% unclassified; of those classified, 22.2% labeled AF
High/low heart rate (outside 50–150 bpm)	99.5% unclassified
Paced rhythms	15.4% unclassified, no trials labeled AF

Clinical Validation

Experiment Design

Apple sponsored a two-part, multicenter study to validate the ability of the ECG app to (1) generate an ECG waveform similar to a Lead I ECG from a standard 12-Lead ECG, and (2) utilize a rhythm classification algorithm to use this single-lead ECG to classify heart rhythms as either SR or AF.

The study's primary endpoints were sensitivity of the rhythm classification algorithm in detecting AF and specificity in detecting SR. An external Institutional Review Board (IRB) approved the protocol, the Informed Consent Form (ICF), and all other relevant materials prior to subject enrollment, and all subjects provided written consent to participate prior to enrollment.

Study participants with known AF and others with no known cardiac rhythm abnormalities were enrolled and asked to record three single-lead ECGs using the ECG app as study staff simultaneously recorded three 12-Lead ECGs using an FDA-cleared clinical device (GE Healthcare CardioSoft ECG device). Participants were given assistance with Apple Watch placement and instructed to keep their arms still, potentially by resting their arms on a table or their leg, and were allowed to practice sample acquisition prior to testing.

To test (1), three independent, certified cardiac technicians overlaid the generated rhythm strips from 140 randomly selected subjects (70 AF, 70 SR) onto the corresponding Lead I strip from the clinical device–generated rhythm strips to visually compare morphology of six consecutive PQRST complexes. Technicians assigned each strip a pass/fail designation based on visually assessed morphological similarity, and were also asked to measure the R amplitude from isoelectric baseline to the nearest millimeter for the first two QRS complexes in both the reference and ECG app–generated rhythm strips and assess the agreement between the two.

For (2), each 12-Lead ECG reference strip was reviewed by three blinded, independent, U.S. board-certified cardiologists who classified the rhythm as either SR, AF, other (anything that was not SR or AF within the heart rate parameters), or unreadable (a diagnosis could not be made as the strip was not adequate for reading). The ECG app algorithm classified the ECG app–generated ECG as SR, AF, unclassifiable, or

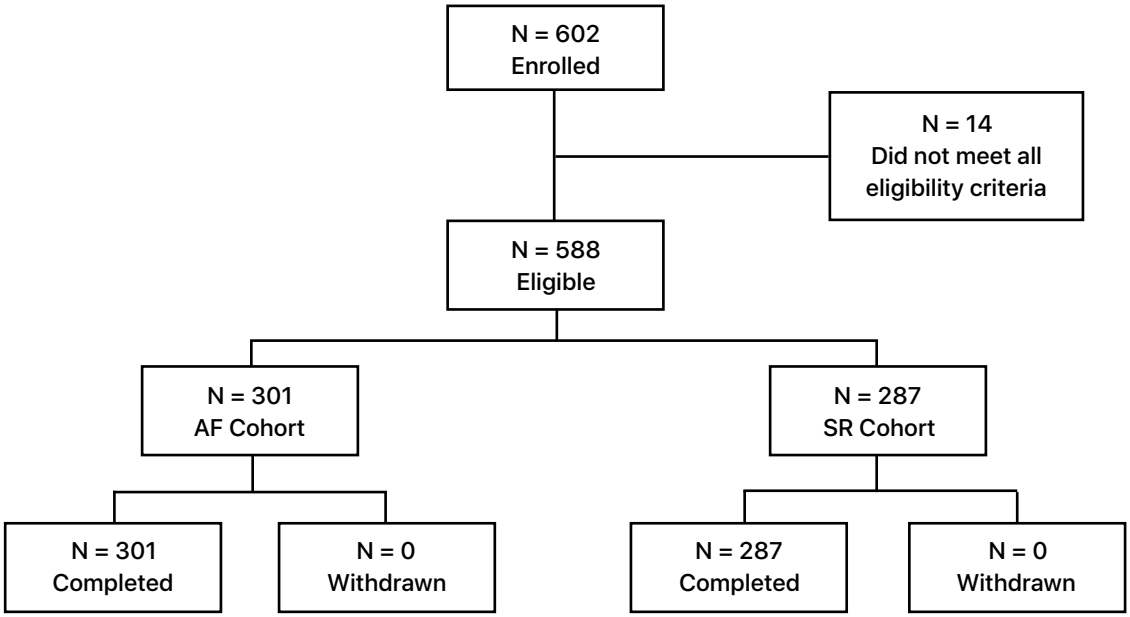
unreadable. The sensitivity and specificity of the ECG app classification of SR and AF (for classifiable ECGs) compared with cardiologist interpretation of the 12-Lead ECG was calculated. One blinded, independent, U.S. board-certified cardiologist was then asked to classify the ECG app–generated rhythm strips according to the same categories.

For primary endpoint analyses, a one-sided exact 97.5% lower confidence bound was computed separately for sensitivity and specificity. If the lower bound for sensitivity exceeded 90%, the null hypothesis, H01, was rejected in favor of the sensitivity exceeding 90%. If the lower bound for specificity exceeded 92%, the null hypothesis, H02, was rejected in favor of the specificity exceeding 92%.

Results

The study enrolled 602 subjects; 588 subjects met eligibility criteria. Of the 588 eligible subjects, 301 subjects with self-reported AF were assigned to the AF Cohort and 287 subjects without self-reported AF were assigned to the SR Cohort. These cohort assignments were used to ensure adequate enrollment only—evaluators were blinded to cohort, and the presence or absence of AF was based solely on the ECG obtained during testing. The 14 subjects who completed the study but were not assigned to an enrollment cohort were ineligible for study participation because of a history of paroxysmal AF without AF on ECG at the time of screening. All eligible subjects completed the study (Figure 2). No adverse events were reported during the study.

Figure 2: Flow Chart of Subject Disposition



Three independent, certified cardiac technicians found visual morphological equivalence between the ECG app waveform and the reference Lead I ECG generated by the standard clinical device for 98.4% of analyzed strips in the AF cohort and 100% in the SR cohort (Table 2). The proportion of overall subjects with a pass rating was 99.2% (lower 97.5% confidence bound, 95.7%). Strips were excluded if six consecutive beats (PQRST complexes) without artifact could not be identified in either set of strips (ECG app or reference).

Table 2. Waveform Comparison

Characteristic	AF Subjects (N = 61)	SR Subjects (N = 65)	Total (N = 126)	Lower Confidence Bound*	p-value**
Number of paired subject strips (ECG app and reference strips) with a pass rating	60	65	125		
Number of readable paired subject strips (ECG app and reference strips)	61	65	126		
Proportion of subject strips with a pass rating	60/61 (98.4%)	65/65 (100%)	125/126 (99.2%)	95.7%	< 0.0001
Number of paired subject strips excluded	8	5	13		

*Lower exact binomial one-sided 97.5% confidence bound for Total. **Test of hypothesis for subject success > 0.8. Abbreviations: AF = Atrial Fibrillation, SR = Sinus Rhythm

For further confirmation that the strips generated by the ECG app and the reference device were similar, blinded cardiologist classification of the ECG app strips was compared with cardiologist classification of the reference strips (Table 3). The percent concordance of the device strip classification with the AF and SR reference results were 100% and 99.1%, respectively. Unreadable strips were not included in this analysis.

Table 3. Classifications Between ECG App and Reference Strips

Characteristic	Total (N = 522)
Final ECG Reference Result = AF	263
Classification of ECG app strip = AF	239/263 (90.9%)
Classification of ECG app strip = SR	0/263 (0.0%)
Classification of ECG app strip = Other	0/263 (0.0%)
Classification of ECG app strip = Unreadable	24/263 (9.1%)
% Concordance with AF Reference Result*	239/239 (100.0%)
Final ECG Reference Result = SR	244
Classification of ECG app strip = AF	0/244 (0.0%)
Classification of ECG app strip = SR	232/244 (95.1%)
Classification of ECG app strip = Other	2/244 (0.8%)
Classification of ECG app strip = Unreadable	10/244 (4.1%)
% Concordance with SR Reference Result*	232/234 (99.1%)

Table 3. Classifications Between ECG App and Reference Strips (continued)

Characteristic	Total (N = 522)
Final ECG Reference Result = Other	15
Classification of ECG app strip = AF	0/15 (0.0%)
Classification of ECG app strip = SR	3/15 (20.0%)
Classification of ECG app strip = Other	12/15 (80.0%)
Classification of ECG app strip = Unreadable	0/15 (0.0%)
% Concordance with Other Reference Result*	12/15 (80.0%)

*Unreadable strips were excluded. Abbreviations: AF = Atrial Fibrillation, SR = Sinus Rhythm

A total of 485 out of 602 paired ECG app and reference rhythm strips were deemed classifiable. The remaining pairs had ECG app and/or reference strips that were deemed unreadable or unclassifiable. Table 4 displays the breakdown among the AF and SR cohorts.

Table 4. ECG App Algorithm Classification and Reference Strip Final Result

ECG App Algorithm Classification	Reference Strip Classification				
	SR	AF	Other	Unreadable	Total
SR	238	4	4	1	247
AF	1	236	2	2	241
Unclassifiable	6	7	6	0	19
Unreadable	18	30	1	0	49
Device Result Not Reported*	32	13	1	0	46
Total	295	290	14	3	602

*Results not reported based on preestablished criteria (such as sync not detected) for all but one subject. Abbreviations: AF = Atrial Fibrillation, SR = Sinus Rhythm

The ECG app algorithm classification achieved a 98.3% sensitivity and 99.6% specificity (Table 5). Expanding analysis to include the 2.8% (7/247) and 2.4% (6/245) of strips categorized as unclassifiable by the device in the AF and SR cohorts, sensitivity was 95.5% (95% CI: 92.2%, 97.8%) and specificity was 97.1% (95% CI: 94.2%, 98.8%). These results met the primary endpoints prespecified in the design of this study. Additionally, 12.2% (68/556) of recordings were inconclusive (either unreadable or unclassifiable) and not classifiable as either SR or AF. When inconclusive recordings were included in the analysis, the ECG app correctly classified SR in 90.5% (238/263) of subjects with SR, and AF in 85.2% (236/277) of subjects with AF. The clinical validation results reflect use in a controlled environment. Real world use of the ECG app may result in a greater number of strips being deemed inconclusive and not classifiable.

Table 5. Sensitivity and Specificity Analysis—Classifiable Strips

Parameter	Value	Lower Confidence Bound*	p-value**
Final ECG Reference Result = AF (n)	240		
ECG App Device Result = AF	236/240 (98.3%)		
ECG App Device Result = SR	4/240 (1.7%)		
Sensitivity	236/240 (98.3%)	95.8%	< 0.0001
Final ECG Reference Result = SR (n)	239		
ECG App Device Result = AF	1/239 (0.4%)		
ECG App Device Result = SR	238/239 (99.6%)		
Specificity	238/239 (99.6%)	97.7%	< 0.0001

*Lower exact binomial one-sided confidence bound.
**Test of hypothesis for sensitivity > 0.9 and specificity > 0.92.
Abbreviations: AF = Atrial Fibrillation, SR = Sinus Rhythm

Conclusions

With the release of watchOS 5.1.2 and Apple Watch Series 4, Apple Watch customers now have access to two optional features enabling detection of irregular heart rhythms. A PPG-enabled algorithm classifies opportunistically collected tachograms in the background, notifying consumers who activate the feature to the presence of an irregular heart rhythm. Apple Watch Series 4 has an electrical heart sensor that, along with the ECG app and algorithm, generates an ECG similar to a single-lead (Lead I) ECG to look for the presence of AF. The proprietary algorithm designed to classify these ECGs as SR, AF, or inconclusive demonstrated > 98% sensitivity and > 99% specificity when compared with ECGs recorded with a reference device and interpreted by independent clinical experts. Consumers wishing to use these features must complete a user experience designed to provide education regarding the interpretation and nondiagnostic nature of these findings and the limitations of the algorithms. Both of these features were granted De Novo classification by the FDA.