Diagnostic performance of single-lead electrocardiograms from the Apple watch and CART ring for cardiac arrhythmias

Andre Briosa e Gala, MD,[1,2,3] Alexander James Sharp, MBBS, BSc,[1,6] David Schramm, MBiochem,[1] Michael Timothy Brian Pope, PhD BM,[3] Milena Leo, PhD,[5] Richard Varini, MBBCh,[1] Abhirup Banerjee, PhD,[6] Kyaw Zaw Win, MD,[7] Manish Kalla, DPhil,[7] John Paisey, DM,[3] Nick Curzen, PhD,[2,3] Timothy Rider Betts, DM[1,4]

From the [1]Department of Cardiology, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom, [2]Faculty of Medicine, University of Southampton, Southampton, United Kingdom, [3]Wessex Cardiac Unit, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom, [4]NIHR Biomedical Research Centre, Oxford, United Kingdom, [5]Queen Alexandra Hospital, Portsmouth, United Kingdom, [6]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom, and [7]Queen Elizabeth Hospital, Birmingham, United Kingdom.

BACKGROUND Wearable devices are widely used for atrial fibrillation (AF) detection, yet most validation studies include only sinus rhythm or AF, likely overestimating diagnostic performance.

OBJECTIVE This multicenter study assessed the performance of automated AF detection and physician interpretation of single-lead electrocardiograms (SL-ECGs) from the Apple Watch and CART Ring.

METHODOLOGY Participants underwent simultaneous 12-lead ECG and SL-ECGs from Apple Watch and CART Ring. Two cardiologists independently adjudicated all ECGs. Apple Watch and CART Ring classified recordings as "AF," "Not AF," or "Unclassified." Diagnostic performance for automated AF detection was evaluated in "worst-case" (all SL-ECGs) and lenient (excluding unclassified SL-ECGs) scenarios. Physician interpretation of SL-ECGs was also compared to 12-lead ECG.

RESULTS Among 483 patients (median age, 66 years; 29% female), 196 (39%) had AF across 3 United Kingdom centers. A total of 2398 ECGs were analyzed. Interobserver variability was excellent (Cohen's kappa: Apple Watch, 0.85; CART Ring, 0.84). In the "worst-case" analysis, CART Ring outperformed Apple Watch (sensitivity, 84.6% vs 69.1%; specificity, 89.9% vs 72.6%). Apple Watch had more unclassified SL-ECGs (20.1%) than CART Ring (1.9%). The lenient analysis showed an improvement in sensitivity (CART Ring, 84.8 %; Apple Watch, 86.4%) and specificity (CART Ring, 91.2%; Apple Watch, 91.7%). Physician interpretation improved diagnostic performance for AF and sinus rhythm but remained limited for other arrhythmias

CONCLUSION Apple Watch missed approximately 1 in 3 episodes of AF and a high number of unclassified SL-ECG. CART Ring demonstrated superior performance. Physician interpretation significantly improved AF diagnosis but remained unreliable for other arrhythmias, emphasizing the need for cautious integration of wearable ECGs into clinical practice.

KEYWORDS Apple watch; CART ring; Atrial fibrillation; Wearable devices; mHealth; Electrocardiogram; Ambulatory monitoring; Digital health

(Heart Rhythm O^2 2025;■:1–10)

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, and it is associated with increased thromboembolic risk, heart failure, and death.[1,2] Conventional AF screening tools, such as 12-lead electrocardiograms (ECGs), Holter monitors, or patches, have inherent limitations in the detection of infrequent arrhythmic episodes because of their brief period of monitoring.[3] This can lead to delays in AF diagnosis and insufficient data to adequately inform effective management strategies.

Over the past decade, a plethora of wearable devices have been equipped with advanced optical sensors that track heart rate with photoplethysmography, generate single-lead ECGs (SL-ECGs), and detect AF. These devices are an appealing and convenient alternative to traditional screening tools and are increasingly being used by patients. Numerous studies have demonstrated their feasibility for AF detection in large

Address reprint requests and correspondence: Andre Briosa e Gala, MD, Oxford University Hospitals NHS Foundation Trust, Headley Way, Oxford, OX3 7BA, UK. E-mail addresses: andre.gala@uhs.nhs.uk; a.briosaegala@gmail.com.

2666-5018/

https://doi.org/10.1016/j.hroo.2025.03.019

KEY FINDINGS

- In patients with a broad range of cardiac rhythms, the CART Ring outperformed the Apple Watch for automated atrial fibrillation detection, with higher sensitivity (84.6% vs 69.1%) and significantly fewer unclassified ECGs (1.9% vs 20.1%).
- The Apple Watch missed approximately 1 in 3 AF episodes, raising concerns regarding its effectiveness as a standalone screening tool.
- Physician interpretation of single-lead ECGs dramatically improved diagnostic performance for both devices, with sensitivity over 94%.
- Both devices performed poorly in detecting non-AF arrhythmias, highlighting a limitation for comprehensive arrhythmia monitoring.
- These findings suggest that hybrid models combining automated detection with physician interpretation could maximize the clinical utility of wearable ECG technologies in real-world practice.

populations.[4,5] However, most validation studies have only included patients with either sinus rhythm or AF, potentially overestimating diagnostic performance by excluding other clinically relevant arrhythmias, which may be important sources of both false-positive and false-negative detections.[6]

The Apple Watch was the first Food and Drug Administration–approved smartwatch to record SL-ECGs. In Apple's clinical validation study with 545 patients, it reported that its Series 6 watch (algorithm version 1.0) had 98% sensitivity and 99% specificity for AF detection.[7] Similarly, the CART Ring is the first CE-marked (Conformité Européenne) ring capable of recording a 30-second SL-ECG to its companion app by placing a finger in its metal casing. The ring is designed to maximize comfort and improve adherence. The clinical validation study of the CART ring SL-ECGs, funded by Sky Labs, Inc (***Gyeonggi-do, South Korea)***, also reported excellent diagnostic metrics for AF detection, with a sensitivity of 99.6%.[8]

Although these validation studies highlight excellent diagnostic performance, their applicability in real-world settings remains uncertain, particularly in patients with a wide range of cardiac rhythms beyond sinus rhythm and AF. In this multicenter study, we evaluate the diagnostic performance of the Apple Watch and CART Ring, focusing on automated AF detection and physician interpretation of SL-ECGs in a cohort of patients with a wide spectrum of arrhythmias.

Methods

Study design and participants

The WEAR-TECH ECG study was a multicenter investigator-initiated diagnostic study conducted at 3 tertiary hospitals in the United Kingdom (Oxford University Hospital, University Hospital Southampton, and Queen Elizabeth Hospital) between December 2021 and December 2022. The study enrolled adult patients (aged ≥18 years) with a documented history of cardiovascular disease in their electronic patient records, defined broadly as any clinically diagnosed disorder affecting the heart or blood vessels. Cardiovascular conditions included, but were not limited to, AF or atrial flutter, ischemic heart disease, heart failure, valvular heart disease, cardiomyopathy, hypertension, and peripheral vascular disease.

Patients were eligible if they were undergoing a 12-lead ECG as part of routine clinical care. We recruited participants in various clinical areas, including coronary care units, inpatient wards, and outpatient clinics, as well as those undergoing elective procedures in the catheterization laboratories. Exclusion criteria was limited to tattoos on the wrists or fingers, which could potentially interfere with the accuracy of device recordings.

All participants provided written informed consent before enrollment. The study complied with the principles of the Declaration of Helsinki and CONSORT guidelines. Ethical approval was granted by the Health Research Authority (ref: 291671), and the study was registered on ClinicalTrials.gov (NCT05298553).

Study aim

The primary objective was to compare the diagnostic performance of the AF detection algorithms of the Apple Watch Series 6 (watchOS 7.0.2, Watch6,1, Algorithm Version 1) and the Skylabs CART Ring using physician-interpreted simultaneous 12-lead ECGs as the reference in patients with known cardiovascular disease. Secondary objectives included: evaluating the diagnostic performance for other atrial arrhythmias (atrial flutter and atrial tachycardia [AT]); assessing physician rhythm interpretation of SL-ECGS compared with 12-lead ECGs; evaluating interobserver agreement between physicians interpreting SL-ECGs; and determining the proportion and distribution of unclassified recordings from each device.

Study procedures

Participants were randomly assigned in a 1:1 fashion to 1 of 2 investigation arms (Supplemental Figure 1). The Robust Randomisation app developed by the Icahn School of Medicine at Mount Sinai was used. Group 1 underwent SL-ECG recording with the CART Ring first, followed by the Apple Watch, and group 2 underwent the Apple Watch recording first, followed by the CART Ring (Supplemental Figure 2).

Each SL-ECG recording was performed simultaneously with a 30-seconds standard 12-lead ECG. If the initial recording was deemed uninterpretable because of artifacts, a second recording was allowed.

The 12-lead ECGs were recorded using an ECG machine with a sweep speed of 25 mm/s and an amplitude of 10 mm/mV on millimetric paper. Recordings from the wearable devices followed their respective instruction manuals, with the

research team providing demonstrations. The Apple Watch generated a 30-second SL-ECG by placing the participant's opposite-hand finger on the watch's metal crown, and the CART Ring generated a 30-second SL-ECG when participants placed a finger on the ring's metal electrode.

After completing the SL-ECG recordings with both devices, participation in the study was concluded, with no follow-up visit required. All SL-ECG recordings from the wearable devices were stored in their respective smartphone app and subsequently exported in PDF format for adjudication. The 12-lead ECG machine also automatically generated and exported recordings in PDF format to the participant's electronic patient records.

ECG adjudication and statistical analysis

Two independent cardiologists adjudicated all SL-ECG recordings and 12-lead ECGs, with any disagreements resolved by a third cardiologist. All ECGs were assigned a rhythm diagnosis, which included AF, sinus rhythm, atrial flutter/AT, ectopy, heart block, ventricular tachycardia, or paced rhythm. Automated algorithms classified SL-ECG recordings into 3 categories: "AF," "Not AF" (labeled as sinus rhythm by the Apple Watch), and "Unclassified." For the Apple Watch, unclassified recordings were further categorized by the device as "Poor recording," "Heart rate above 120 beats/min," "Heart rate below 50 beats/min," and "Inconclusive," according to its internal criteria. ECG quality was graded on a 3-point scale: 1 (high, confident diagnosis with minimal artifact), 2 (intermediate, diagnosis possible despite artifact), and 3 (uninterpretable, recording quality too poor for diagnostic interpretation).

To assess the diagnostic performance of automated algorithms for identifying AF, SL-ECGs were compared with 12-lead ECGs, which served as the reference. To evaluate the impact of unclassified SL-ECGs on diagnostic performance, 2 analytic approaches were employed. The first, termed "worst-case" scenario analysis, reflects a real-world scenario in which unclassified ECGs cannot be further interpreted. In this approach, unclassified SL-ECGs by the wearable devices were treated as false positives if the corresponding 12-lead ECG showed no AF, and as false negatives if the corresponding 12-lead ECG confirmed AF. This approach provided a lower-bound estimate of diagnostic performance. The second, lenient analysis, reflected ideal conditions by excluding unclassified SL-ECGs, aiming to isolate the diagnostic performance of the algorithms under optimal circumstances. To assess performance for atrial arrhythmias, a detection was considered correct if either the algorithm or the physician interpretation of SL-ECGs identified any atrial arrhythmia (including AF, atrial flutter, or AT). This approach acknowledges the clinical significance of detecting any form of atrial arrhythmia, even if the specific subtype was not categorized correctly. Additionally, we evaluated the diagnostic performance of physician interpretation of SL-ECGs for all cardiac rhythms.

Continuous variables were presented as means with standard deviations or medians with interquartile ranges, according to their distribution. Categorical variables were expressed as frequencies and percentages. Outcome measures for each device tested included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy. Comparison of diagnostic performance metrics was performed by using McNemar's test for paired proportions. For subgroup analyses, the Fisher's exact test was used. Interobserver agreement between cardiologists was formally quantified using Cohen's kappa coefficient, in which values >0.80 were indicative of excellent agreement. Statistical tests were 2-tailed and performed using R (version 4.1.3), with statistical significance set at $P < .05$.

Results

A total of 508 patients with cardiovascular disease were enrolled across 3 centers over a 12-month period. After excluding 10 patients with suboptimal 12-lead ECG recordings and 15 patients with ventricular pacing, 483 patients and 2398 ECGs were included in the final analysis. The median age of participants was 66 years (interquartile range, 57–75), and 29% were female. A history of AF was reported in 60.8% of patient, and 11.8% had a history of atrial flutter or AT. The presenting rhythm on the 12-lead ECG was AF in 194 (40%) of patients. Baseline characteristics are listed in Table 1.

Physician adjudicators achieved 91.7% agreement for Apple Watch and 89.7% for CART Ring SL-ECG interpretations. Interobserver variability was excellent: Cohen's kappa values of 0.85 for the Apple Watch and 0.84 for the CART Ring. No significant difference ($P = .70$) in the diagnostic quality of SL-ECGs was observed, with 78.1% and 79.3% of recordings categorized as a high diagnostic quality (Supplemental Table 1).

Diagnostic performance for atrial fibrillation detection

Automated AF detection algorithms demonstrated notable differences in performance (Table 2, Figure 1). In the "worst-case" scenario analysis, which included all SL-ECGs, the Apple Watch had a sensitivity of 69.1% (95% confidence interval [CI], 62.3–75.2) and a specificity of 72.6% (95% CI, 67.1–77.4). In contrast, the CART Ring had superior performance, with a sensitivity of 84.6% (95% CI, 78.9–89.0) and a specificity of 89.9% (95% CI, 85.9–92.9), both statistically significant ($P < .01$).

The lenient analysis, with the exclusion of unclassified SL-ECGs (97 in the Watch group and 9 in the Ring group), was associated with improved performance for both devices. The Apple Watch showed an increase in sensitivity by 15.1% and in specificity by 19.1%. In this lenient analysis, there was no significant difference in sensitivity between the 2 devices (84.8% vs 86.4%). False-positive AF detections in the Apple Watch were largely attributable to atrial flutter/AT (62.5%) and sinus rhythm (31.3%), and the CART Ring showed

Table 1 Baseline characteristics and demographics for the whole cohort, according to the presenting rhythm on the 12-lead ECG

Characteristics	Total (n = 483)	AF (n = 194)	Not AF (n = 283)	*P*
Age, median (Q1–Q3)	66 (57–75)	67 (60–74)	65 (54–75)	.041
Female sex, n (%)	139 (29%)	57 (29%)	57 (29%)	.9
BMI, median (Q1–Q3)	27.8 (24.9–32.5)	28.9 (25.7–32.9)	27.5 (24.5–31.6)	.06
Ethnicity, n (%)				.6
Caucasian	469 (97%)	192 (98%)	277 (96%)	
Co-morbidities, n (%)				
Ischemic heart disease	113 (22.6%)	20 (10%)	88 (31%)	<.001
Myocardial infarction	68 (13.7%)	12 (6.2%)	55 (19%)	<.001
Atrial fibrillation	292 (60.8%)	182 (93%)	110 (38%)	<.001
Atrial tachycardia/flutter	59 (11.8%)	22 (11%)	37 (12%)	.9
Congestive cardiac failure	101 (21%)	53 (27%)	48 (17%)	.008
Hypertension	153 (32%)	69 (35%)	84 (29%)	.2
Valvular heart disease	86 (18%)	46 (24%)	40 (14%)	.009
Cardiac surgery	50 (10%)	22 (11%)	28 (9.7%)	.7
Stroke/TIA	29 (6.0%)	20 (10%)	9 (3.1%)	.02
Diabetes mellitus	34 (7.0%)	16 (8.2%)	18 (6.3%)	.5
Anticoagulation, n (%)				
DOACS	250 (51.7%)	110 (36%)	140 (49%)	<.001
Warfarin	20 (4.0%)	12 (6.2%)	8 (2.6%)	.07
Rate control drugs, n (%)				.006
Beta-blocker	245 (50.7%)	140 (72%)	105 (36.%)	
Calcium-channel blocker	20 (4.0%)	12 (6.2%)	8 (2.6%)	
Digoxin	12 (2.8%)	6 (3.1%)	6 (2.2%)	
Antiarrhythmic drugs, n (%)				.016
Amiodarone	42 (8.7%)	21 (7.3%)	21 (11%)	
Dronedarone	2 (0.4%)	—	2 (0.7%)	
Flecainide	23 (4.8%)	4 (2.1%)	19 (9.8%)	
Propafenone	2 (0.4%)	1 (0.5%)	1 (0.3%)	
Sotalol	19 (3.9%)	12 (6.2%)	7 (2.4%)	

AF = atrial fibrillation; BMI = body mass index; DOACS = direct oral anticoagulants; ECG = electrocardiogram; TIA = transient ischemic attack.

similar trends, with atrial flutter/AT (60%) and more contributions from sinus rhythm with ectopy (13.3%) (Supplemental Table 2).

Physician interpretation of SL-ECGs markedly improved AF detection. Interpretation of SL-ECGs from the Apple Watch demonstrated slightly higher sensitivity of 95.4% (95% CI, 91.4–97.5) compared with the CART Ring (94.3%; 95% CI, 90.1–96.8), whereas the specificity of the Apple Watch (89.6%; 95% CI, 85.6–92.6) closely matched that of the CART Ring (88.9%; 95% CI, 84.8–92.0).

Unclassified ECGs

Unclassified SL-ECGs accounted for 20.1% (97) of the SL-ECGs recorded by the Apple Watch and only 1.9% (9) of those recorded by the CART Ring (Table 3). Among the unclassified recordings from the Apple Watch, 62% were categorized as "Inconclusive," with 72% originating from ECGs demonstrating either AF or sinus rhythm. Twenty-four percent of the unclassified SL-ECGs resulted from heart rates outside the diagnostic range (<50 beats/min or >120 beats/min). In contrast, the CART Ring had a markedly lower rate of unclassified (1.9%) SL-ECGs, with most poor-quality recordings seen in AF (44%) and sinus rhythm (33%). Physician interpretation correctly identified the rhythm on 76.2% (74) of unclassified Apple Watch ECGs and 66.7% (6) of unclassified CART Ring SL-ECGs.

Diagnostic performance for atrial arrhythmias

Among 242 patients with atrial arrhythmias, 194 had AF and 48 had atrial flutter or AT. In the "worst-case" scenario analysis for atrial arrhythmias, the CART Ring exhibited superior sensitivity at 77.2% (95% CI, 69.5–80.3) compared with 59.3% (95% CI, 53.0–65.2) for the Apple Watch (Figure 2, Table S3). This trend mirrors the CART Ring's advantage observed in automated AF detection. The diagnostic performance of automatic detection for atrial arrhythmias was lower than for AF because the algorithm is designed to detect an irregularly irregular rhythm, and atrial tachycardia can be regular or have limited irregularity. When unclassified ECGs were excluded (lenient analysis), it resulted in a marked increase in sensitivity of the Apple Watch to 75.8% (95% CI, 69.2–81.3) and specificity to 96.9% (95% CI, 93.5–98.6), comparable to the CART Ring. Physician interpretation of SL-ECGs improved diagnostic performance considerably for both the Apple Watch and CART Ring, with sensitivities of 96.2% and 90.8%, respectively.

Of the 48 atrial flutter/AT ECGs, 24 had variable AV nodal conduction and an irregular ventricular rate, of which the Apple Watch identified 9 (38%) as AF compared with 13 (54%) using the CART Ring (Table 4). With a regular ventricular rate, the Apple Watch classified only 1 (4.2%) as AF, 11 (46%) as unclassified, and 12 (50%) as sinus rhythm. The CART Ring identified 5 (21%) regular rate cases as AF, with the remainder classified as sinus rhythm.

Table 2 Diagnostic performance of the automatic detection and physician-interpretation of single-lead ECG for AF using the 12-lead ECG as gold standard

Device	AF detection	N	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Apple Watch	Algorithm (all ECGs)	483	69.1 (62.3–75.2)	72.6 (67.1–77.4)	62.9 (56.2–69.1)	77.7 (72.4–82.3)	71.0 (66.8–74.9)
	Algorithm (unclassified excluded)	386	84.8 (78.4–89.6)	91.7% (87.4–94.6)	87.6 (81.4–91.9)	89.7 (85.1–93.0)	88.9 (85.3–91.6)
	Physician (all ECGs)	483	95.4 (91.4–97.5)	89.6 (85.6–92.6)	86.0 (80.8–90.0)	96.6 (93.7–98.2)	91.9 (89.2–94.0)
CART Ring	Algorithm (all SL-ECG)	483	84.6 (78.9–89.0)	89.9 (85.9–92.9)	85.1 (79.4–89.4)	89.6 (85.6–92.6)	85.1 (84.7–90.4)
	Algorithm (unclassified excluded)	474	86.4 (80.8–90.5)	91.2 (87.3–93.9)	86.8 (81.3–90.9)	90.8 (86.9–93.7)	89.2 (86.1–91.7)
	Physician (all SL-ECG)	483	94.3 (90.1–96.8)	88.9 (84.8–92.0)	85.1 (79.7–89.3)	95.9 (92.8–97.7)	91.1 (88.2–93.3)

AF = atrial fibrillation; CI = confidence intervals; ECG = electrocardiogram; NPV = negative predictive value; PPV = positive predictive value; SL-ECG = single-lead electrocardiogram.

Physician interpretation for other cardiac rhythms

Figure 3 demonstrates the distribution of true positives, false positives, and false negatives across all cardiac rhythms. Sensitivity for sinus rhythm was excellent, 92.1% (95% CI, 87.6–95.0) for the Apple Watch and 87.3% (95% CI, 82.2–91.2) for the CART Ring, with specificity of more than 92% for both devices. However, performance for other arrhythmias, such as atrial flutter and ectopy, was poor, and these were frequently misclassified (Supplemental Figure 3). The CART Ring had double the number of SL-ECGs classified as uninterpretable compared with the Apple Watch (17 [3.5%] vs 7 [1.5%]; $P = .032$).

Impact of SL-ECG quality

The diagnostic quality of SL-ECG recordings significantly influenced the performance of automated AF detection and physician interpretation of cardiac rhythms for both devices (Supplemental Table 5, Supplemental Figure 4). The CART Ring demonstrated superior automated detection compared with the Apple Watch for high-quality SL-ECGs (90.2% vs 76.6%; $P < .001$). With intermediate-quality SL-ECGs, the Apple Watch's accuracy declined substantially to 51.7%, whereas the CART Ring maintained higher accuracy at 87.7% ($P < .001$). Notably, 51% of unclassified SL-ECGs from the Apple Watch were of high quality, compared with 32% from the CART Ring. Physician interpretation of cardiac rhythms remained robust across both devices, with an accuracy of over 80% and only modest declines observed between high- and intermediate-quality recordings.

Discussion

This multicenter study offers valuable insight into the diagnostic performance of the Apple Watch and CART devices in a "real-word" cohort of patients with different cardiac rhythms. Our findings reveal that sensitivity and specificity for automated AF detection were lower than previously reported, particularly for the Apple Watch, which missed 1 in 3 episodes of AF under the "worst-case" analysis. In contrast, the CART Ring demonstrated greater diagnostic reliability, with a significantly lower unclassified ECG rate (1.9% vs 20.1%), likely reflecting a more robust algorithm processing and signal quality.

A key limitation of the Apple Watch was the large proportion of unclassified SL-ECGs, which is consistent with previous peer-reviewed studies (19%–28%) but represents a significant challenge for patients relying on these devices for AF screening and monitoring.[9,10] In this study, we used algorithm version 1.0 of the Apple Watch, which has a heart rate limit of 120 beats/min for AF detection. Apple has since released algorithm version 2, which increased the limit to 150 beats/min. Newer Apple Watch models may also incorporate improved sensors, which may enhance signal quality. In Apple's internal validation studies, which have not been published in peer-reviewed journals, the rate of unclassified ECGs decreased from 12.2% with version 1.0 to 7.1% with

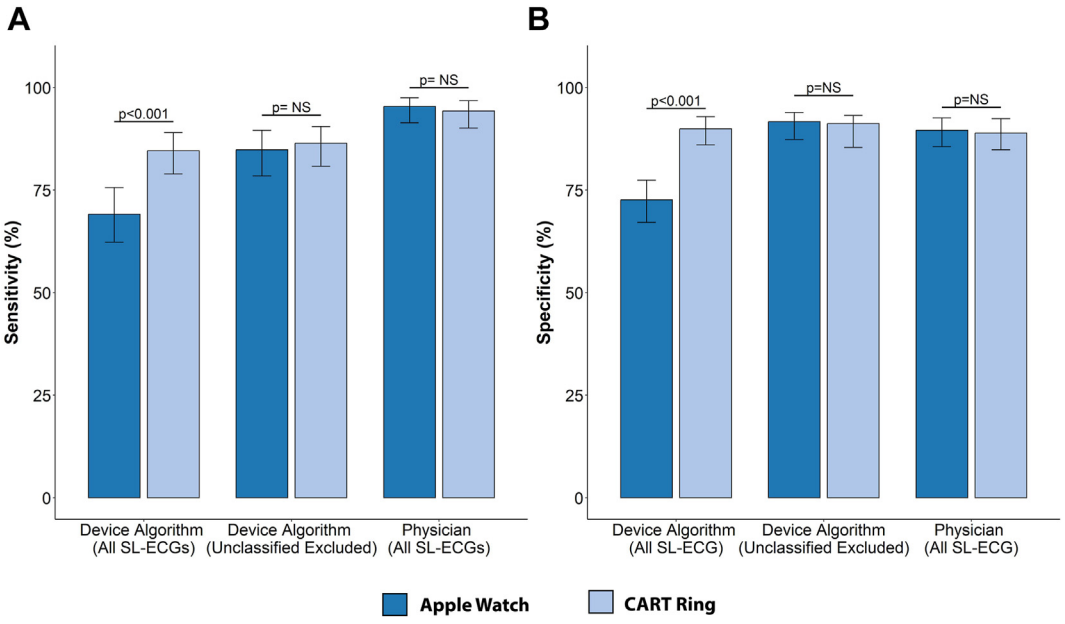

Figure 1 Diagnostic performance of the Apple Watch and CART Ring for AF detection compared with a 12-lead ECG. **A:** Sensitivity and **B:** specificity across 3 analytic categories: the worst-case scenario analysis (including all SL-ECGs), the lenient analysis (excluding unclassified ECGs), and physician interpretation of SL-ECGs.

version 2.0.[7,11] However, the rates of unclassified ECGs in independent studies have consistently been higher, underscoring the need for external validation of these devices.

A recent study involving 247 participants and testing 5 smartwatches found that, despite multiple software updates and increased user experience, there was only a modest reduction in unclassified SL-ECGs over time.[12] Pepplinkhuizen et al[13] demonstrated that repeat recordings after an unclassified SL-ECG on the Apple Watch reduced the overall number of unclassified readings; however, this came at the cost of reduced accuracy and increased false-positive rates. In contrast, the CART Ring had a considerably lower rate (1.9%) (Supplemental Figure 3). Among the 97 SL-ECGs unclassified by Apple Watch, the CART Ring correctly identified the rhythm in 81% of cases. These findings suggest that superior signal quality from its ring sensor and its cloud-based algorithm mitigates some of the limitations of on-device processing, allowing for more precise rhythm classification.

The clinical utility of wearable devices is determined not only by their sensitivity, but their specificity and PPV also should be taken into account. For AF screening, high specificity is paramount to minimize false-positive detection, which can lead to unnecessary anxiety and overuse of health care resources, and may even result in inappropriate anticoagulation.[14,15] The combination of low specificity and a high number of unclassified SL-ECGs is concerning if the Apple Watch were used as a screening tool. For ongoing AF management, high sensitivity is desired to avoid missing clinically significant recurrences that may influence management strategies.

Physician interpretation of SL-ECGs for AF and sinus rhythm was excellent for both devices, reinforcing the 2020 European Society of Cardiology guidelines, which, for the

Table 3 Classification of unclassified ECGs from the Apple Watch and CART Ring compared with corresponding rhythms identified on 12-lead ECGs

	Apple Watch				CART Ring
12- Lead ECG	HR < 50 beats/min (n = 16)	HR > 120 beats/min (n = 8)	Inconclusive (n = 61)	Poor recording (n = 12)	Poor recording (n = 9)
Atrial fibrillation	1 (6.2%)	3 (38%)	28 (45%)	5 (42%)	4 (44%)
Sinus rhythm	9 (56%)	1 (12%)	16 (27%)	6 (50%)	3 (33%)
Atrial flutter/AT	1 (6.2%)	4 (50%)	11 (18%)	—	2 (22%)
Sinus rhythm with ectopy	1 (6.2%)	—	4 (6.7%)	1 (8.3%)	—
AF with complete heart block	—	—	—	—	—
Complete heart block	3 (19%)	—	1 (1.7%)	—	—
Ventricular tachycardia	—	—	1 (1.7%)		—
Junctional rhythm	—	—	—	—	—

AT = atrial tachycardia; CHB = complete heart block.

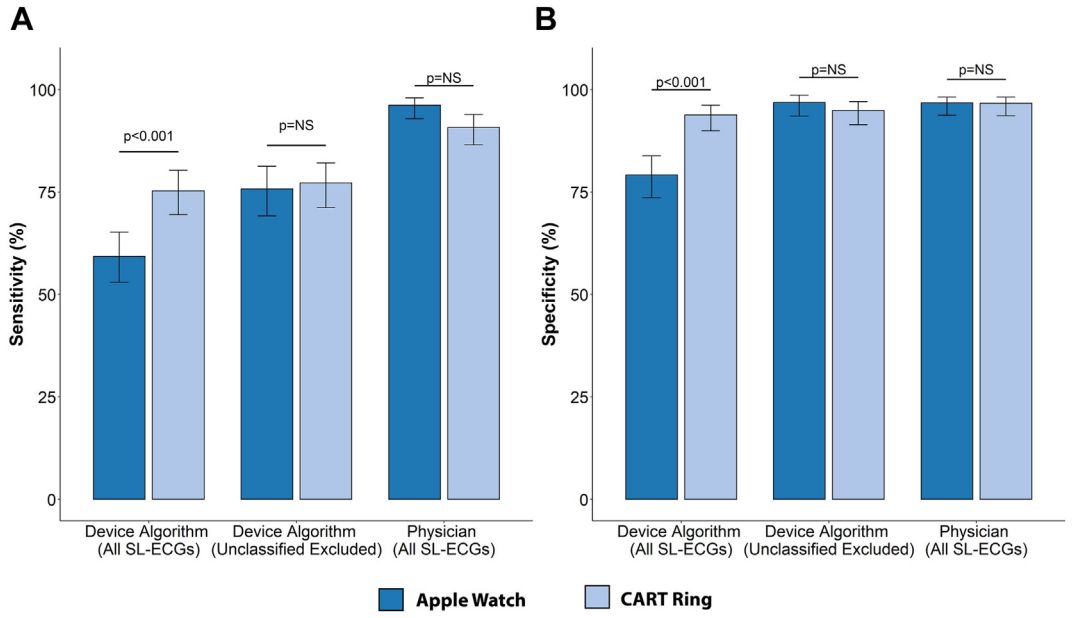

Figure 2 Diagnostic performance of the Apple Watch and CART Ring for atrial arrhythmias compared with a 12-lead ECG. **A:** Sensitivity and **B:** specificity across 3 analytic categories: the worst-case scenario analysis (including all SL-ECGs), the lenient analysis (excluding unclassified ECGs), and physician interpretation of SL-ECGs.

first time, recognized physician-interpreted SL-ECGs from wearable devices as evidence for the diagnosis of AF.[16] However, despite the large proportion of high-quality SL-ECGs in this study, physician interpretation of other rhythms was suboptimal. The heavily filtered nature of SL-ECGs has the effect of obscuring key features, such as P-waves, that are essential for the diagnosis of atrial flutter/AT and heart blocks. Atrial and ventricular ectopy, however, were comparatively easier to identify, because their detection mostly relies on QRS timing and morphology. Although several case reports have showcased the potential utility of SL-ECGs in detecting other cardiac rhythms, our findings stress the need for cautious interpretation, and management decision should ideally be corroborated by a 12-lead ECG.[17,18]

Atrial flutter and AT detection remain a major limitation of wearable ECGs. In this study, physician interpretation of atrial flutter and AT had poor sensitivity across both devices: 37.5% for the Apple Watch and 22.9% for the CART Ring. These challenges have significant implications for the clinical use of wearable devices in AF management. First, freedom

Table 4 Automated detection and physician interpretation for atrial flutter/AT with regular and irregular block using all SL-ECG recordings ("worst-case" scenario analysis)

Device	Detection method	Atrial flutter/AT with irregular block (n = 24)	Atrial flutter/AT with regular block (n = 24)	*P*
Apple Watch	**Automated detection**			**.012**
	• Atrial fibrillation	9 (38%)	1 (4.2%)	
	• Sinus rhythm	10 (42%)	12 (50%)	
	• Unclassified	5 (21%)	11 (46%)	
	Physician interpretation			**.002**
	• Atrial fibrillation	18 (75%)	6 (25%)	
	• Atrial flutter/AT	5 (21%)	13 (54%)	
	• Sinus rhythm	1 (4.2%)	5 (21%)	
CART Ring	**Automated detection**			**.036**
	• Atrial fibrillation	13 (54%)	5 (21%)	
	• Sinus rhythm	10 (42%)	18 (75%)	
	• Unclassified	1 (4.2%)	1 (4.2%)	
	Physician Interpretation			**<.001**
	• Atrial fibrillation	21 (88%)	5 (21%)	
	• Atrial flutter/AT	—	11 (46%)	
	• Sinus rhythm	1 (4.2%)	5 (21%)	
	• Sinus rhythm with ectopy			
	• Uninterpretable			

AT = atrial tachycardia; SL-ECG = single-lead electrocardiogram.

A

Physician Interpretation Apple Watch SL-ECG	12-Lead ECGs: AF (n=194)	AT/Flutter (n=48)	Sinus rhythm (n=214)	Sinus rhythm & ectopy (n=17)	AF CHB (n=2)	CHB (n=5)	VT (n=1)	Junctional rhythm (n=1)
AF	**185 (95.4%)**	24 (50%)	5 (2.4%)	1 (5.8%)	—	—	—	—
AT/Flutter		**18 (37.5%)**	2 (0.9%)		—	—	—	—
Sinus rhythm	1 (0.5%)	6 (12.5%)	**197 (92.1%)**	6 (35.4%)	1 (50%)	4 (80%)	—	—
Sinus rhythm & ectopy	2 (1%)	—	9 (4.2%)	**10 (58.8%)**	—	—	—	1 (100%)
AF CHB	—	—	—	—	**1 (50%)**	—	—	—
CHB	—	—	—	—	—	**1 (20%)**	—	—
VT	—	—	—	—	—	—	—	—
Junctional rhythm	—	—	—	—	—	—	—	—
Uninterpretable	5 (2.6%)	—	1 (0.5%)	—	—	—	1 (100%)	—

B

Physician Interpretation CART ring SL-ECG	12-Lead ECGs: AF (n=194)	AT/Flutter (n=48)	Sinus rhythm (n=213)	Sinus rhythm & ectopy (n=17)	AF CHB (n=2)	CHB (n=5)	VT (n=1)	Junctional rhythm (n=1)
AF	**183 (94.3%)**	26 (54.2%)	5 (2.3%)	1 (5.9%)	—	—	—	—
AT/Flutter	—	**11 (22.9%)**	1 (0.5%)	—	—	—	1 (100%)	—
Sinus rhythm	1 (0.5%)	6 (12.5%)	**186 (87.3%)**	7 (41.2%)	2 (100%)	4 (80%)	—	1 (100%)
Sinus rhythm & ectopy	3 (1.5%)	4 (8.3%)	11 (5.2%)	**9 (52.9%)**	—	1 (20%)	—	—
AF CHB	—	—	—	—	—	—	—	—
CHB	—	—	—	—	—	—	—	—
VT	—	—	—	—	—	—	—	—
Junctional rhythm	—	—	—	—	—	—	—	—
Uninterpretable	7 (3.6%)	1 (2.1%)	9 (4.2%)	—	—	—	—	—

Figure 3 Physician interpretation of single-lead ECGs from the Apple Watch (**A**) and CART Ring (**B**) compared with rhythms identified on 12-lead ECGs. AF = atrial fibrillation; AT = atrial tachycardia; CHB = complete heart block; VT = ventricular tachycardia.

from atrial arrhythmias is a key outcome after catheter ablation, and wearable devices are increasingly leveraged to monitor for arrhythmia recurrence.[19,20] The inability to reliably detect atrial flutter or AT undermines the utility of wearables in monitoring treatment success. Second, recurrence of atrial arrhythmias often dictates subsequent catheter ablation strategies, making accurate differentiation between AF and other atrial arrhythmias critical.[20]

Moreover, poor rhythm discrimination is a challenge for new rhythm-guided anticoagulation strategies, such as "pill-in-the-pocket" oral anticoagulation, in which therapy is initiated during and shortly after an atrial arrhythmia.

The REACT-AF trial (NCT05836987), a multicentre, prospective, randomized study of over 5,300 patients, is underway to compare rhythm-guided anticoagulation using the Apple Watch compared with continuous anticoagulation. Approximately one-fifth of patients with AF also have concomitant atrial flutter; these episodes may go undetected and do not trigger appropriate anticoagulation, leaving patients at risk of thromboembolic events.[21,22]

As wearable technologies evolve, machine learning and artificial intelligence (AI) will likely continue to further improve their diagnostic metrics, learning through repeated use and training on increasingly large datasets.[23] These innovations may potentially mitigate some of the limitations of current technologies identified in this study, such as the high rate of unclassified SL-ECGs and the inability to accurately identify non-AF arrhythmias. Similarly, progress has been made with implantable cardiac monitors, where the dual-state AI algorithm of the Reveal LINQ II (Medtronic, Minneapolis, MN) led to an 88.2% reduction in false-positive AF alerts.[24,25] Comparable findings have also been reported with a convolutional neural network classifying AF episodes from the Jot Dx (Abbott, Abbott Park, IL).[26]

The strengths of this study are the inclusion of a broad range of arrhythmias, the use of simultaneous testing with the gold-standard 12-lead ECG, limiting the number of SL-ECG attempts, creating a rigorous and clinically relevant testing environment. Prior validation studies have potentially overestimated diagnostic performance by focusing solely on AF and sinus rhythm.[27] Ranine et al,[10] in a similar study that included arrhythmias beyond AF and sinus rhythm, showed that automated AF detection had 69% sensitivity and 81% specificity, aligning with our findings.

This study had several limitations. Our population was older adults with established cardiovascular disease, limiting the generalizability of findings to younger or asymptomatic individuals who increasingly use wearable devices. The higher prevalence of atrial arrhythmias in our study cohort impacts on performance metrics, such as PPV and NPV, and therefore these must be interpreted with caution. Our results are limited to Apple Watch algorithm version 1.0, and results may differ from improved algorithm versions, which require further testing. The CART Ring software version tested has not been disclosed and may not reflect newer iterations. We also limited the number of SL-ECGs recordings to 2 per device, which may account for the high unclassified rate of SL-ECGs. Our unclassified rates are, however, comparable to those reported in studies without these restrictions, and reflects a more realistic assessment of real-world usability. The SL-ECGs in this study were recorded under physician supervision, likely contributing to good signal quality. In everyday use, in which recordings are done independently by users, variability in signal quality may be greater, potentially resulting in even more unclassified ECGs. Finally, manual adjudication by 2 independent cardiologists, although strengthening clinical validation, introduces the possibility of human error despite the high level of agreement we observed between adjudicators.

Conclusion

The CART Ring's automated AF detection algorithm outperformed the Apple Watch, with greater sensitivity (84.6% vs 69.1%) and a significantly lower rate of unclassified SL-ECGs (1.9% vs 20.1%). Physician interpretation improved diagnostic performance for both AF and sinus rhythm; however, it remained poor for non-AF arrhythmias, such as atrial flutter and atrial tachycardia.

From a clinical perspective, these findings emphasize the limitations of relying solely on automated AF detection algorithms for AF diagnosis and guiding management decisions. The variability in sensitivity and specificity highlights the importance of caution when incorporating wearable devices' automated detection into patient care without adequate physician interpretation. Hybrid models that combine automated analysis supplemented with physician interpretation should be prioritized in clinical workflows to maximize the diagnostic utility and reliability of wearable technologies.

Funding: Funding was provided by Oxford Cardiac Rhythm Managament Research, and Skylabs loaned the CART rings used in this study. A. B. is supported by the Royal Society University Research Fellowship (Grant No. URF\R1\221314).

Disclosures: A. B. G., A. J. S., D. S., M. T. B. P., M. L., R. V., K. Z. W., M. K., A. B., J. R. P., and T. R. B. declare no CI for this study. N. C. reports grant funding from Haemonetics, HeartFlow, Boston Scientific & Beckman Coulter; speaker fees/consultancy from Abbott, HeartFlow, Edwards.

Authorship: All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent: Participants provided written informed consent for data collection and analysis.

Ethics Statement: Ethical approval was granted by the Health Research Authority (ref: 291671), and the study was registered on ClinicalTrials.gov (NCT05298553).

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hroo.2025.03.019.

References

1. Li H, Song X, Liang Y, et al. Global, regional, and national burden of disease study of atrial fibrillation/flutter, 1990–2019: results from a global burden of disease study, 2019. BMC Public Health 2022;22:2015.
2. Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 ESC guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): developed by the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC), with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). Eur Heart J 2024;45:3314–3414.
3. Charitos EI, Stierle U, Ziegler PD, et al. A comprehensive evaluation of rhythm monitoring strategies for the detection of atrial fibrillation recurrence: insights from 647 continuously monitored patients and implications for monitoring after therapeutic interventions. Circulation 2012;126:806–814.
4. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909–1917.
5. Lubitz SA, Faranesh AZ, Selvaggi C, et al. Detection of atrial fibrillation in a large population using wearable devices: the fitbit heart study. Circulation 2022;146:1415–1424.

6. Linz D, Gawalko M, Betz K, et al. Atrial fibrillation: epidemiology, screening and digital health. Lancet Reg Health Eur 2024;37:100786.
7. Apple Inc. Using Apple Watch for Arrhythmia Detection. December 2020. https://www.apple.com/healthcare/docs/site/Apple_Watch_Arrhythmia_Detection.pdf. Accessed November 26, 2024.
8. Kwon S, Hong J, Choi EK, et al. Detection of atrial fibrillation using a ring-type wearable device (CardioTracker) and deep learning analysis of photoplethysmography signals: prospective observational proof-of-concept study. J Med Internet Res 2020;22(5).
9. Mannhart D, Lischer M, Knecht S, et al. Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL wearable study. JACC Clin Electrophysiol 2023;9:232–242.
10. Racine HP, Strik M, van der Zande J, et al. Role of coexisting ECG anomalies in the accuracy of smartwatch ECG detection of atrial fibrillation. Can J Cardiol 2022;38:1709–1712.
11. Apple INC. ECG app 2.0 Algorithms Clinical Validation Study, https://cdn.clinicaltrials.gov/large-docs/81/NCT04247581/Prot_000.pdf.
12. Isenegger C, Mannhart D, Arnet R, et al. Accuracy of smartwatches for atrial fibrillation detection over time: insights from the Basel wearable study. JACC Clin Electrophysiol 2024;10(12):2735–2737.
13. Pepplinkhuizen S, Hoeksema WF, van der Stuijt W, et al. Accuracy and clinical relevance of the single-lead Apple Watch electrocardiogram to identify atrial fibrillation. Cardiovasc Digit Health J 2022;3:S17–S22.
14. Force UPST. Screening for atrial fibrillation: US Preventive Services Task Force Recommendation Statement. JAMA 2022;327:360–367.
15. Rosman L, Lampert R, Zhuo S, et al. Wearable devices, health care use, and psychological well-being in patients with atrial fibrillation. J Am Heart Assoc 2024;13:e033750.
16. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021;42(40):4194.
17. Galgut O, Le Page P, Mitchell ARJ. Watch for tachycardia. Int J Arrhythmia 2022;23:30.
18. Yerasi C, O'Donoghue S, Satler LF, et al. Apple Watch detecting high-grade block after transcatheter aortic valve implantation. Eur Heart J 2019;41:1096. 1096.
19. Kalman JM, Al-Kaisey AM, Parameswaran R, et al. Impact of early vs. delayed atrial fibrillation catheter ablation on atrial arrhythmia recurrences. Eur Heart J 2023;44:2447–2454.
20. Calkins H, Hindricks G, Cappato R, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm 2017;14:e275–e444.
21. Aro A, Eyob-Fesseha H, Haukka J, et al. How often atrial flutter and atrial fibrillation coexist? Results from a large nationwide ECG-based study. Europace 2023;25(Suppl 1):euad122.037.
22. Waldo AL, Feld GK. Inter-relationships of atrial fibrillation and atrial flutter: mechanisms and clinical implications. J Am Coll Cardiol 2008;51:779–786.
23. Li Y-H, Li Y-L, Wei M-Y, et al. Innovation and challenges of artificial intelligence technology in personalized healthcare. Sci Rep 2024;14:18994.
24. Radtke AP, King K, Mei K, et al. PO-03-046 Accelerating improvements to arrythmia classifiers with AI-curated datasets. Heart Rhythm 2023;20:S484.
25. Singh J, Radtke A, Rosemas S, et al. Impact of enhanced artificial intelligence on clinic burden from false alerts of insertable cardiac monitors. Circulation 2023;148:A18672. A18672.
26. Betts TR, Gardner RS, Quartieri F, et al. PO-678-08 Neural network model for automatic discrimination of atrial fibrillation episodes detected by an insertable cardiac monitor. Heart Rhythm 2022;19:S352–S353.
27. Zarak MS, Khan SA, Majeed H, et al. Systematic review of validation studies for the use of wearable smartwatches in the screening of atrial fibrillation. Int J Arrhythmia 2024;25:11.