Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation

Kevin Rajakariar 1,1 Anoop N Koshy 1,2 Jithin K Sajeev,1 Sachin Nair,3 Louise Roberts,1,3 Andrew W Teh1,3

ABSTRACT
Objective  The AliveCor KardiaBand (KB) is an Food and Drug Administration-approved smartwatch-based cardiac rhythm monitor that records a lead-intelligent ECG (IECG). Despite the appeal of wearable integrated ECG devices, there is a paucity of data evaluating their accuracy in diagnosing atrial fibrillation (AF). We evaluated whether a smartwatch-based device for AF detection is an accurate tool for diagnosing AF when compared with 12-lead ECG.

Methods  A prospective, multi-centre, validation study was conducted in an inpatient hospital setting. The KB paired with a smartwatch, generated an automated diagnosis of AF or sinus rhythm (SR). This was compared with a 12-lead ECG performed immediately after IECG tracing. Where an unclassified or no-analysis tracing was generated, repeat IECG was performed.

Results  439 ECGs (IECGs (n=239) and 12-lead ECG (n=200)) were recorded in 200 patients (AF: n=38; SR: n=162) from three tertiary centres. Sensitivity and specificity using KB was 94.4% and 81.9% respectively, with a positive predictive value of 54.8% and negative predictive value of 98.4%. Agreement between 12-lead ECG and KB diagnosis was moderate when unclassified tracings were included (κ=0.60, 95% CI 0.47 to 0.72). Combining the automated device diagnosis with blinded electrophysiologists (EP) interpretation of unclassified tracings improved overall agreement (EP1: κ=0.76, 95% CI 0.65 to 0.87; EP2: κ=0.74, 95% CI 0.63 to 0.86).

Conclusion  The KB demonstrated moderate diagnostic accuracy when compared with a 12-lead ECG. Combining the automated device diagnosis with EP interpretation of unclassified tracings yielded improved accuracy. However, even with future improvements in automated algorithms, physician involvement will likely remain an essential component when exploring the utility of these devices for arrhythmia screening.


INTRODUCTION
Wearable arrhythmia detection devices are becoming increasingly popular, with sales set to double by 2022.1-4 These devices obtain physiological data using a simple user interface to instantaneously provide a diagnosis and transmit information to a receiving cardiologist for prompt review. With consumer interest driving the rapid expansion of portable healthcare, this ultimately culminated in the US Food and Drug Administration (FDA) clearing several technologies including the AliveCor KardiaBand (KB), which is a strap for the Apple watch (Apple, Cupertino, California, USA) with an inbuilt electrode that generates a lead-intelligent ECG (IECG) and provides an automated diagnosis of either atrial fibrillation (AF) or sinus rhythm (SR). If proven to be accurate, these devices have the potential to be used for population-level screening of AF with subsequent prescription of oral anticoagulation to prevent stroke. However, before population screening can be effective, the accuracy of the diagnostic test requires rigorous validation. While there is high clinician interest and consumer expectation of diagnostic precision from these devices, there is a paucity of data evaluating their accuracy in arrhythmia diagnosis.5 In addition, although such devices are freely available to the consumer, there is insufficient data from clinical trials to guide clinicians on how to interpret their data and integrate it into the decision-making process. We aimed to evaluate the accuracy of the KB as a detection tool for AF in an unselected hospital cohort.

METHODS
A prospective, multi-centre validation study was performed across three tertiary university hospitals, where consecutive patients ≥18 years of age admitted to the medical, cardiac or intensive care ward were invited to participate. Patients with cardiac implantable electronic devices, those unable to independently use the device, or in contact isolation were excluded from the study. Written informed consent was obtained from all subjects. Baseline characteristics including demographic data, cardiac risk factors and medications were obtained.

The KB strap was attached to an Apple Watch and paired with an iPhone 6 smartphone (Apple, Cupertino, California, USA) using the AliveCor Kardia application V5.0.2 (AliveCor, Mountain View, California, USA). The device obtains a 30-second continuous lead-1 recording that can be viewed in real-time on the iPhone and is remotely transmitted to a secure server (US Health Insurance Portability and Accountability standards compliant) for storage and subsequent clinician analysis. The KB analysis then displays a diagnosis of either (a) possible AF;
b) normal SR; c) no-analysis or d) unclassified. The no-analysis reading is displayed when there is excessive artefact or interference for the KB algorithm to determine a diagnosis. An unclassified reading is shown when, in the absence of recording artefact, the device is unable to determine whether the trace is AF or SR. Bradycardia was defined as heart rate <50 beats per minute (bpm) and tachycardia ≥100bpm, and when detected resulted in an unclassified reading. An ECG trace was generated by wearing the Apple Watch on the wrist, and placing the thumb of the opposite hand onto the electrode located on the KB strap for 30 seconds. Where an unclassified or no-analysis reading was displayed, a second KB tracing was obtained and used as the final diagnosis (figure 1). All patients were supervised and instructed in a standardised manner, and the initial recorded iECG was used with no test recordings performed prior.

Patients were screened for inclusion on the wards. After recruitment, patients were consented and an iECG tracing was obtained. A 12-lead ECG was performed immediately following the KB trace, for assessment of KB accuracy and to minimise the chance of rhythm variance between tracings. There were no opportunities for patients to perform additional iECG tracings. KB tracings were only obtained during initial recruitment and were not taken during symptomatic episodes or at any alternative time. The ECG was interpreted and diagnosis subsequently determined by a cardiologist (ANK) and used as the reference standard against the KB device. The ECG automated diagnosis was not used. Sensitivity, specificity, predictive value, likelihood ratios were calculated using 2×2 contingency tables. Accuracy (A) was estimated by determining the percentage of true positive (TP) and true negative (TN) results compared with all results (R) when analysing KB accuracy by assessing unclassified diagnoses as incorrect, AF was categorised as a false negative and SR as a false positive. Although some arrhythmias were correctly given an unclassified diagnosis, they were considered incorrect to conservatively measure KB accuracy and in a real-world scenario, all unclassified diagnoses would require clinician interpretation. All unclassified tracings were compiled for diagnostic assessment by two independent blinded cardiac electrophysiologists (EP1: JKS, EP2: AWT). Due to the artefactual nature of no-analysis readings, consecutive iECGs generating this diagnosis were not included in the statistical analysis. Overall agreement between the KB, electrophysiologists, and 12-lead ECG was analysed using Cohen’s kappa coefficient (κ). All analyses were two-tailed, and p values of <0.05 were considered statistically significant. Statistical analyses were performed using Stata V.13/MP.

**Patient and public involvement statement**

The patients and public were not involved in the creation of the study design, recruitment or statistical analysis. Patients were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

**RESULTS**

A total of 218 consecutive patients were recruited over 6 months, with 200 patients (56.5% male, age 67±16 years) eligible for inclusion into the study. Overall, 439 ECGs were recorded (iECGs (n=239) and 12-lead ECG (n=200)). This included 38 (19%) patients in AF, and 162 in SR based on the interpreted 12-lead ECG diagnosis (determined by ANK). Eighteen patients were excluded as they did not fulfil inclusion criteria (implanted cardiac device (n=8), significant tremor (n=6), contact isolation (n=4)). The baseline clinical characteristics are summarised in table 1. Patients with AF were more likely to be older with more comorbid conditions.

A summary of the iECG results is shown in figure 2. Overall, 161 (80.5%) AF and SR diagnoses were obtained from the initial iECG (figures 3 and 4). Fourteen tracings (7%) were recorded as no-analysis and 25 (12.5%) as unclassified readings (figures 5 and 6). A second iECG performed (n=39) recorded a diagnosis in 14 patients (7%) with the remainder being unclassified (8%) or no-analysis (4.5%). In total, 175 (87.5%) AF and SR diagnoses were obtained using up to two iECGs from the patient cohort. On review of unclassified diagnoses, seven (3.5%) were given the diagnosis in the setting of sinus tachycardia, three (1.5%) contained premature atrial complexes, three (1.5%) were sinus rhythm with first degree atrioventricular block, and three (1.5%) iECGs had unclear reasons for remaining unclassified. No tracings were given an unclassified diagnosis due to bradycardia. All no-analysis tracings were in the setting of artefact (figure 7), including two (1%) patients unable to complete the tracing due to Parkinsonian tremor.

The iECG automated algorithm demonstrated an overall sensitivity of 94.4% when categorising unclassified readings as incorrect, and improved to 95.4% when incorporating tracings which
Table 1  Patient demographics and clinical characteristics

<table>
<thead>
<tr>
<th></th>
<th>Normal sinus rhythm (n=162)</th>
<th>Atrial fibrillation (n=38)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years (mean±SD)</td>
<td>64±17</td>
<td>76±11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>64%</td>
<td>52%</td>
<td>0.64</td>
</tr>
<tr>
<td>Ischaemic heart disease</td>
<td>32%</td>
<td>50%</td>
<td>0.09</td>
</tr>
<tr>
<td>Hypertension</td>
<td>51%</td>
<td>50%</td>
<td>0.36</td>
</tr>
<tr>
<td>Heart failure</td>
<td>13%</td>
<td>44%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>20%</td>
<td>25%</td>
<td>0.53</td>
</tr>
<tr>
<td>Stroke/TIA</td>
<td>16%</td>
<td>17%</td>
<td>0.93</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>10%</td>
<td>14%</td>
<td>0.07</td>
</tr>
<tr>
<td>Known atrial flutter/fibrillation</td>
<td>9%</td>
<td>95%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of palpitations</td>
<td>26%</td>
<td>79%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean CHA2DS2VASc score</td>
<td>2</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>Antiarrhythmic medications (any)</td>
<td>6%</td>
<td>44%</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Flecainide 1% 0%
Beta-blocker 22% 100%
Amiodarone/sotalol 5% 14%
Calcium channel blocker 1% 0%
Digoxin 1% 14%

Acute embolic event

Wangfarin 0% 0%
Apixaban 4% 57%
Rivaroxaban 4% 21%
Dabigatran 0% 14%
Low molecular weight heparin 0% 7%

P values<0.05 demonstrate a significant difference between patients with normal sinus rhythm and atrial fibrillation. Patients in atrial fibrillation were older, with a higher incidence of palpitations and heart failure.

Figure 2  Summary of iECG results—a total of 39 (19.5%) iECGs were uninterpretable and subsequently repeated, of which 14 (7%) generated a diagnosis. iECG, intelligent ECG.

SR. The KB specificity was 81.9%, which increased to 90.7% when excluding unclassified readings due to a reduction in false positive diagnoses. The positive and negative predictive values for the KB (excluding unclassified diagnoses) were 72.3% and 98.4% respectively. Agreement between the single lead KB diagnosis and 12-lead ECG cardiologist diagnosis when including unclassified tracings (marked as incorrect) was moderate (κ=0.60, 95%CI 0.47 to 0.72, p<0.001), however improved when incorporating appropriately diagnosed unclassified readings (κ=0.71, 95%CI 0.62 to 0.77, p<0.001) and excluding unclassified readings altogether (κ=0.76, 95%CI 0.67 to 0.88, p<0.001).

Review of unclassified tracings by two electrophysiologists blinded to the 12-lead ECG diagnosis revealed an overall accuracy of 93.8% (EP1: 1 false positive, 0 false negatives) and 87.5% (EP2: 2 false positives, 0 false negatives) respectively. When incorporating EP diagnoses of unclassified readings with the KB automated analysis, we demonstrate improved positive predictive value (70.8% and 69.4% for EP1 and EP2 respectively) and improved agreement with the 12-lead ECG cardiologist diagnosis (EP1: κ=0.76, 95%CI 0.63 to 0.87, p<0.001; EP2: κ=0.74, 95%CI 0.63 to 0.86, p<0.001). Interobserver agreement between EPs was strong (κ=0.98, 95%CI 0.91 to 1.00, p<0.001). A summary of results with clinician interpretation is outlined in table 2.
The majority of patients with AF had a pre-existing history of the condition. However, there were two new diagnoses of AF, confirmed on both 12-lead ECG cardiologist interpretation and iECG automated diagnosis. Both patients described a history of palpitations, but were asymptomatic during the time of the AF recordings.

DISCUSSION

This study evaluated the accuracy of a novel smartwatch-based ECG monitor for diagnosing AF in an unselected high-risk inpatient population. When a KB automated diagnosis was proffered, it demonstrated high sensitivity for diagnosing AF. A significant number of false positives were noted for both device automated diagnosis of AF and clinician interpretation of unclassified readings. Although overall KB accuracy was moderate when categorising unclassified tracings as incorrect, incorporating the KB automated diagnosis with EP interpretation of only unclassified tracings yielded improved diagnostic accuracy.

The initial iECG also demonstrated a 19.5% rate of undiagnosed tracings, which nearly halved once a repeat iECG was performed, consistent with similar studies using AliveCor technology. The exact reasons for the relatively high rates of no analysis and unclassified tracings is unclear. However, recruitment of older, hospitalised patients could have affected the ability for patients to record the tracings without generation of artefact that may have altered the accuracy of the device algorithm (figure 5). While many of these tracings were due to electrical noise during the period of iECG acquisition, some of our unclassified diagnoses were appropriate in certain non-AF arrhythmias such as sinus bradycardia or tachycardia (figure 6). Such rhythms are more likely to occur in a hospitalised cohort and therefore cautious interpretation is required when analysing the device rate of unclassified diagnoses. In addition, where an unclassified or no-analysis diagnosis is specified, a repeat iECG should be performed to help differentiate from artefactual causes for the given diagnosis.

There is increasing interest in the use of smartphone-based technology in AF detection with numerous published feasibility and screening studies. However, multiple thorough validation studies are required to confirm their accuracy before they can be recommended for population level use, with the ultimate goal being accurate early detection of AF leading to prevention of stroke. Since the recent FDA approval and release of the smartwatch-based ECG devices, only two previous studies have reported on its accuracy. However, both studies experienced high dropout rates, with one trial conducted in a preselected patient cohort undergoing cardioversion, limiting its generalisability. This is particularly pertinent given that pretest probability of AF significantly affects testing parameters with prior studies reporting a positive predictive value as low as 5% for AF in an unselected patient population. In our study, in an unselected hospital patient population, the KB algorithm yielded high sensitivity with a significant number of unclassified readings; this would be expected to improve over time with rapid developments in AF detection algorithms including the use of deep neural networks. In addition, given the significant proportion of false positives, using a diagnosis of AF from the device to guide commencement of treatment such as oral anticoagulants for stroke prophylaxis cannot be recommended at this stage. With increasing adoption of wearable arrhythmia devices, establishing the accuracy of these technologies is crucial and needs to occur before clinicians can begin to even consider whether subsequent investigations and therapies should be instituted.

With rapid developments in arrhythmia detection technology, there are numerous advantages of smartwatch linked devices compared with their smartphone counterparts. Improved accessibility of smartwatch-based devices allows the consumer the
ability to perform tracings throughout the day ensuring greater opportunities for detection of paroxysmal AF. Moreover, the incorporation of AliveCor SmartRhythm technology into the KB uses inbuilt smartwatch photoplethysmography to provide an opportunity for continuous cardiac monitoring. This technology employs a deep neural network to correlate acute fluctuations in heart rate with physical activity, and recent studies have demonstrated excellent accuracy compared with other wearable devices.\(^{10}\) Any discordance between the two will prompt the user to acquire an ECG.\(^ {11}\) Similar technology has been incorporated to the newly released Apple smartwatches, however the accuracy of these technologies has not been published. AliveCor has subsequently released the KardiaMobile 6L, an arrhythmia detection device with the ability to create a hand-held 6-lead iECG using three electrodes. This technology may be complementary to the validation studies performed using KB with a potential for providing more detailed electrocardiographic data. With further clinical validation studies of these newer devices, clinicians will be better equipped to understand how to incorporate data from these technologies into their practice.

A chief concern highlighted with AF detection platforms are the associated costs involved with clinician interpretation of iECG tracings. In the largest community study using the AliveCor Kardia monitor, the cost for a new diagnosis of AF exceeded $10,000, largely driven by expenditure associated with commercial over-reads.\(^ {12,13}\) Importantly, consumers are not limited in the amount of KB tracings which can be performed, which could lead to increased false positive diagnoses, unclassified readings and clinicians being overwhelmed by the volume of data requiring oversight. In a pilot study, we proposed a pathway whereby limiting clinician interpretation of iECG tracings to unclassified readings could be a more economically viable strategy for arrhythmia screening when using smartphone-based ECG technology.\(^ {14,15}\) In the present study, incorporating the KB automated diagnosis with EP interpretation of unclassified tracings alone demonstrated improved accuracy, which is comparable with results of previous studies employing clinician interpretation of all tracings.\(^ {20,21}\) In a high-risk population, combining KB diagnoses with clinician over-read of unclassified tracings alone may provide a more cost-effective solution than clinician interpretation of all iECG tracings.\(^ {10}\)

Our proposed work-flow of limiting clinician traces to ‘unclassified’ traces could be challenging in the real-world. Ongoing consumer engagement with these devices may generate increasing numbers of these traces that require clinician over-read. Notwithstanding, this may be more feasible than manual over-read of all ECG tracings. Our findings shed light on the feasibility and accuracy of smartwatch-based arrhythmia detection in a hospitalised cohort. However, large scale cohort studies are needed to confirm our findings and the cost-effectiveness of this strategy in an all-comer community-dwelling population. Collaboration among key stakeholders including industry, clinicians and consumers is needed to generate actionable health policy that can leverage the use of wearable device-generated ECG data in improving care provision.

**Implications**

Wearables can facilitate affordable large-scale arrhythmia screening at a population level. However, there are several barriers to overcome if such technology is to become the contemporary standard in arrhythmia detection. Current iterations of single-lead iECG devices are heavily reliant on symptoms of palpitations to trigger opportunistic screening, and consequently can overlook patients with paroxysmal arrhythmias.\(^ {22}\) While alternative wearables using photoplethysmography (PPG) can provide continuous screening, accuracy is currently affected by motion artefact, ectopic beats, peripheral vascular disease, poor skin contact and limited battery life.\(^ {23}\) In addition, the potential for continuous monitoring is limited by the battery life with current device iterations requiring daily overnight charging. A simplistic user interface will also be required to ensure those at the highest risk of AF—the elderly—will be able to adapt to such technology to not only obtain iECGs, but to interpret its findings and identify when further clinician input is required. Newer devices that can combine PPG with accelerometer data, superior battery life and smarter arrhythmia detection algorithms may overcome the current constraints that limit wearables from becoming the next affordable frontier in arrhythmia screening. With increasing uptake of wearable devices with ECG capabilities, there will be an exponential rise in positive iECG readings. This has potential to culminate in both user anxiety and consumption of valuable health resources for downstream diagnostic testing.\(^ {19}\) Therefore, algorithms designed to streamline clinician over-read of tracings are also required if population level arrhythmia screening is to occur. Finally, long-term outcomes for patients treated for subclinical or device-detected AF is unknown; further studies are required to determine the risk of thromboembolic stroke in this subgroup and the efficacy of treatment with oral anticoagulation.\(^ {24}\)

We acknowledge certain limitations of this study. First, our cohort had an AF prevalence rate of 19%, which is significantly higher than the 2%–5% population estimates of AF.\(^ {25}\) A higher prevalence of AF in our population could overestimate the calculated predictive values in our study. As such, it would be essential to validate these results in an all-comer outpatient population with a lower pretest probability of AF. Second, it remains unclear whether the frequency of unclassified and uninterpretable tracings is likely to be higher when the device is used at home by patients who may lack the knowledge and training to amend their technique to improve iECG acquisition. This, in turn, could
increase both unnecessary patient anxiety and the number of iECGs that necessitate clinician review. In our study, all patients were supervised and provided instructions on technique, which is unlikely to occur in a real-world scenario. Furthermore, even if the KB is established as an accurate tool for diagnosing AF, the threshold at which to treat AF in otherwise asymptomatic individuals remains unknown. Lastly, only two EPs assessed iECGs in this study. It remains unclear if these findings are generalisable to other physicians and whether prior clinician experience with iECGs affects accuracy.

CONCLUSION
In an unselected hospital patient cohort, the KB demonstrated moderate diagnostic accuracy when compared with a 12-lead ECG, which improved when uninterpretable traces were excluded. Combining the automated device diagnosis with EP interpretation of unclassified tracings improved both positive predictive value and overall accuracy. However, even with future improvements in automated algorithms, this suggests that physician involvement will likely remain an essential component when exploring the utility of these devices for arrhythmia screening.

What is already known on this subject?
► Wearable arrhythmia detection devices are becoming increasingly popular, with sales set to double in the next 3 years.
► Despite the growing uptake of wearables in the community, there is a paucity of data evaluating the accuracy of these devices.

What might this study add?
► In a prospective study of 200 patients, the KardiaBand demonstrated moderate diagnostic accuracy compared with a 12-lead ECG when unclassified traces were included.

How might this impact on clinical practice?
► Combining the automated device diagnosis with clinician interpretation of unclassified tracings yielded improved accuracy.
► Even with future improvements in automated algorithms, physician involvement will likely remain an essential component when exploring the utility of these devices for arrhythmia screening.

Twitter Anoop N Koshy @DrAnoop_Koshy

Acknowledgements We thank the physicians, nurses and patients involved in the study.

Contributors KR and ANK equally contributed to the initial manuscript draft, recruitment and data analysis. SN was involved in data collection. JKS and LR contributed to data analysis and draft revision. AWJ reviewed the data, edited and approved the final submitted draft.

Funding This work was supported by the Eastern Health Foundation Research Grant [EHFRG2017_029]. The sponsor had no role in study design, collection, analysis, interpretation of data and in the decision to submit the article for publication. Dr. Koshy is supported by scholarships from the National Health and Medical Research Council of Australia, National Heart Foundation and The Royal Australasian College of Physicians.

Competing interests None declared.

Patient consent for publication Not required.

REFERENCES

Ethics approval Institutional ethics review board approval of the study protocol (LRSS-2016).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. Data are available upon reasonable request.

ORCID iDs
Kevin Rajakariar http://orcid.org/0000-0002-3886-8214
Anoop N Koshy http://orcid.org/0000-0002-8741-9631

Twitter Anoop N Koshy @DrAnoop_Koshy

Acknowledgements We thank the physicians, nurses and patients involved in the study.

Contributors KR and ANK equally contributed to the initial manuscript draft, recruitment and data analysis. SN was involved in data collection. JKS and LR contributed to data analysis and draft revision. AWJ reviewed the data, edited and approved the final submitted draft.

Funding This work was supported by the Eastern Health Foundation Research Grant [EHFRG2017_029]. The sponsor had no role in study design, collection, analysis, interpretation of data and in the decision to submit the article for publication. Dr. Koshy is supported by scholarships from the National Health and Medical Research Council of Australia, National Heart Foundation and The Royal Australasian College of Physicians.

Competing interests None declared.

Patient consent for publication Not required.