Evaluating the Apple Watch for Atrial Fibrillation Detection: Diagnostic Accuracy, Clinical Utility, and User Perceptions

Asha H Golveo

Embry-Riddle Aeronautical University

Dr. Samuel Benavides

March 16, 2025

Abstract

Atrial fibrillation (AFib) is a prevalent cardiac arrhythmia associated with significant risks of stroke, heart failure, and mortality. Early detection is crucial for effective management and prevention of complications. Consumer-grade wearable devices, such as the Apple Watch, have emerged as potential tools for AFib detection through photoplethysmography (PPG) and single-lead electrocardiogram (ECG) technology. This study evaluates the diagnostic accuracy of Apple Watch's AFib detection feature, its impact on healthcare utilization, and user perceptions. A mixed-methods approach will be employed, integrating quantitative analyses of diagnostic accuracy, healthcare utilization, and device reliability with qualitative assessments of patient and healthcare provider experiences. Sensitivity and specificity of the Apple Watch ECG will be compared to standard 12-lead ECG data from clinical trials such as the Apple Heart Study and Heartline Study. The research will also explore demographic factors affecting detection accuracy, including age, skin tone, and technical literacy. Findings from this study will contribute to the growing body of literature on digital health technologies, assessing their feasibility for integration into clinical practice.

Introduction

Atrial fibrillation (AFib) is a cardiac arrhythmia associated with significant risks of stroke, heart failure, and increased mortality. By 2030, an estimated 2.6 million people in the United States will have AFib, and a concerning number of cases remain undetected until serious thromboembolic events occur (Mohamoud et al., 2024). With asymptomatic AFib significantly elevating the risk of stroke, early recognition of AFib is critical. Wearable devices, particularly smartwatches like the Apple Watch, have emerged as a promising tool to improve detection for heart rhythm monitoring.

Photoplethysmography (PPG), the technology that the Apple Watch utilizes for AFib detection, intermittently measures heart rate and rhythm by detecting changes in blood flow. Newer models of the Apple Watch have also incorporated a single-lead electrocardiogram (ECG) feature, enabling on-demand 30-second recordings similar to a Lead 1 ECG, which received 510(k) clearance from the FDA for users 22 years and older (Mohamoud et al., 2024). Recent studies have demonstrated the feasibility of the Apple Watch's PPG to identify irregular pulses, and a meta-analysis of 11 studies indicated high diagnostic accuracy of the Apple Watch ECG for detecting AFib. While these findings suggest the potential of the Apple Watch in AFib detection, the need for more extensive clinical validation is needed. In addition, high risk populations of AFib are infrequently the target demographic of Apple Watches. Additional research is necessary to understand how Apple Watch’s AFib detection will affect clinical utility, cost-effectiveness, and impact on healthcare providers.

This research proposal seeks to address the efficacy of the Apple Watch in detecting atrial fibrillation, its perceived reliability by patients and healthcare providers, and its impact on

clinical utilization. Understanding patient perceptions, and identifying the factors that contribute to the reliability and effective integration of the Apple Watch into clinical practice, is essential in realizing the full potential of wearable technology in improving early diagnosis and management of atrial fibrillation. This research will contribute valuable insights into the real-world impact of consumer-grade wearable ECG devices on patient care and healthcare systems. This study is designed to assess the hypothesis that the Apple Watch ECG and PPG-based AFib detection system demonstrates high sensitivity and specificity in identifying atrial fibrillation compared to a standard 12-lead ECG, making it a reliable tool for early AFib detection.

Literature Review

The Potential of Wearable Devices for AFib Detection

Consumer-grade wearable devices, including smartwatches like the Apple Watch, have emerged as a promising tool for monitoring heart rhythms. Wearable devices utilize technologies like photoplethysmography (PPG) and single-lead electrocardiograms (ECGs) to detect irregular heart rhythms suggestive of AFib. A systematic review and meta-analysis by Prasitlumkum et al. (2021) investigated the diagnostic accuracy of smart devices like smartphones and smartwatches for detecting AFib. The authors analyzed numerous studies comparing these devices to standard electrocardiograms, finding high sensitivity and specificity for both smartphones and smartwatches in detecting AFib.

A study by Elbey et al. (2021) compared smartwatches to medical-grade devices for atrial fibrillation detection and found that smartwatches were “non-inferior to composite ECG monitoring strategies”. In addition, some smartwatches offer the potential for more continuous, passive monitoring compared to intermittent traditional methods like standard ECG or pulse

checks. Some studies suggest that smartwatches using PPG technology can achieve acceptable sensitivity and high specificity for AFib detection, especially in screening populations. However, a study by Wasserlauf et al. (2023) found that sensitivity may be lower for populations with known AFib. Smartwatches have many strengths in AFib detection and monitoring as opposed to traditional ECG strategies, but are currently limited by factors like battery life and the need for stillness during readings (Vyas et al., 2024).

Diagnostic Accuracy of the Apple Watch ECG

Several studies have evaluated the diagnostic performance of the Apple Watch ECG in detecting AFib by comparing it to the standard 12-lead ECG. Shahid et al. (2025) conducted a systematic review and diagnostic test accuracy (DTA) meta-analysis, synthesizing data from eleven studies involving over 4,000 participants to determine the pooled sensitivity and specificity of the Apple Watch's ECG feature. The study conducted a comprehensive literature search on Medline/PubMed (Ovid), Embase (Ovid), and the Cochrane Central Register of Controlled Trials from their inception to March 25, 2024. The search terms included variations of "apple smartwatch," "ECG," and "atrial fibrillation". The inclusion criteria for the meta-analysis focused on studies involving human subjects, evaluating the Apple Watch against 12-lead ECG, diagnosing AFib, and providing estimates of sensitivity and specificity with 95% confidence intervals. The study’s findings suggest that the Apple Watch demonstrates high accuracy in identifying AFib, with the pooled sensitivity at 94.8% (95% CI: 91.7% to 96.8%; I2 ¼ 67%) and specificity at 95% (95% CI: 88.6% to 97.8%; I2 ¼ 88%). The authors did note significant heterogeneity among the included studies, acknowledging the need for further research to validate these results in broader populations, and assess the clinical implications.

A study by Heo et al. (2024) recruited 30 participants from a heart failure clinic to assess the feasibility of monitoring patients with high risk of AFib using an Apple Watch. When participants received an abnormal pulse notification from the Apple Watch, the ECG report generated by the Apple Watch was emailed to a team of cardiologists for review. This study focused on individuals diagnosed with heart problems, including patients with a higher risk of atrial fibrillation, thus allowing them to track which patients sought additional care from their medical providers. While participants found the Apple Watch easy to use, usability was correlated with age, education, and income, highlighting areas for improvement to ensure equitable access. The findings of this study support the potential of the Apple Watch in AFib detection and monitoring, while underscoring the necessity of future research to improve the accessibility and usability of this feature to groups directly impacted by AFib.

Large-Scale Assessments and Clinical Trials

The Apple Heart Study was a large scale, single-arm prospective site-less digital trial conducted to evaluate the accuracy of Apple Watch's algorithm in identifying irregular heart rhythms suggestive of AFib. The study was conducted by Stanford University involving 419,297 participants, with the findings showing 0.52% received notifications of irregular pulses, and 34% of those subsequently showing AFib on ECG (Garcia et al., 2022). Furthermore, the Apple Heart Study demonstrated the feasibility of using a smartwatch for large-scale data collection and participant-reported outcomes via electronic surveys.

The Heartline Study is an ongoing study designed to determine if the Apple Watch's irregular rhythm notification and ECG capabilities can decrease the time to AFib diagnosis, thus facilitating treatment adherence (Dhruva et al., 2021). The randomized controlled trial is a collaboration between Apple and Johnson & Johnson, which aims to address the impact of

personal digital devices on quality of life and healthcare utilization. While the use of wearable devices for health information continues to increase, few studies have assessed their impact on important clinical outcomes, independent of device manufacturers. Prior research has shown that increased screening for AFib can lead to higher rates of AF detection using various cardiac monitoring devices. However, these increased detection rates were not consistently associated with a significant reduction in clinical outcomes (Harxhi et al., 2023).

Broader Impact of Wearable Technologies in Health Research

A scoping review performed by Huhn et al. (2022) aimed to overview and categorize current research, with the focus on affordable, consumer-grade wearables for health research from a population health perspective. This review analyzed 179 studies published between 2013 and 2020 that used noninvasive wearables worn on the wrist, arm, hip, and chest to measure vital signs. The review identified various applications of wearable technology in health research, including diagnostics and screening, disease monitoring, prognosis, forecasting, risk stratification, explorative analysis of big data sets, method evaluation, and feasibility studies. The most frequently emphasized strengths of wearables in these studies were accuracy and reliability, often supported by validation studies and clinical certifications. However, the review also noted shortcomings such as inaccuracies in measurements and technical issues.

Data access policies, and the transparency of algorithms used by wearable companies, are important considerations for researchers, as these can impact data analysis and comparability across studies. In addition, the growing use of wearables in clinical practice and health diagnosis necessitate considerations of equity and fairness in the adoption and regulation of these devices. An integrated approach considering ethical, legal, and social factors is necessary for the

responsible development and inclusion of wearable technology in healthcare programs and research (Canali et al., 2022).

Conclusion

The literature indicates a growing interest in the potential of consumer-grade wearable devices in detecting atrial fibrillation, particularly the Apple Watch. The ubiquity of smart devices suggests that their cost-effectiveness in medical screening could be tremendous, with studies showing encouraging diagnostic accuracy for the Apple Watch ECG compared to standard ECG. The use of wearable devices can extend to various applications in health research, including monitoring, prediction, and method evaluation. Ongoing large-scale clinical trials aim to determine the real-world impact of these devices on clinical outcomes and healthcare utilization. However, factors such as methodological limitations in accuracy studies, the need for validation in diverse populations, data management considerations, and understanding the perspectives of both patients and healthcare providers remain critical areas for further research and integration of this technology into clinical practice. In addition, challenges related to data quality, interoperability, and ethical considerations need to be addressed to fully realize their potential in digital health.

Data Treatment

To address the research problem of Apple Watch's AFib detection accuracy, patients' and healthcare providers' perception of its efficacy, and factors that influence its reliability and clinical utilization, the research would require using a mixed-methods approach. Data collection and analysis would involve both quantitative and qualitative data, with thematic analysis utilized to identify common patterns in attitudes, experiences, and concerns.

Quantitative data will be collected on the diagnostic accuracy of AFib detection of the Apple Watch, as well as data on its technical reliability. Diagnostic accuracy data will focus on the sensitivity and specificity of the Apple Watch ECG in detecting AFib, ideally compared to a standard 12-lead ECG. It is critical that data collection identifies false positives, false negatives, true positives, and true negatives. Furthermore, collecting data on inconclusive tracings and poor recordings are essential to understanding the factors that affect sensor performance, such as skin perfusion, skin tones, or tattoos. Studies have shown that wearable devices utilizing PPG with green light signals may have reduced accuracy in people with dark skin tones, a phenomenon that is underreported and under researched, and affects a large population of wearable device users (Colvonen et al., 2020) In addition to skin tone, understanding the effect of tattoos, sweat, hair, and activity levels on PPG accuracy, is essential as more of these devices are approved by the FDA for clinical research. Usage data can be collected from Apple Watch ECG recordings by patients, tracking the frequency of irregular rhythm notifications received, including the time to diagnosis following use of the Apple Watch. Studies such as The Heartline Study are evaluating the ability of the Apple Watch to decrease the time to AFib diagnosis after patients receive irregular rhythm notifications (Gibson et al., 2023).

Patient demographics such as age, sex, and pre-existing conditions (including history of AFib) will be collected to understand the influence those factors may have on the reliability and utilization of the device. Whereas some studies specifically focus on patients with or without a history of AFib, this study will include both types of patients. The objective is to understand the impact of the AFib feature on the demographic that has greater need for AFib detection, while understanding the impact of the feature on the demographic that has greater adoption of wearable technology. Healthcare utilization data will be collected by gathering information on healthcare

visits (primary care, cardiology, emergency department), hospitalizations, diagnostic testing (ECGs, heart rhythm monitors), and interventions (cardioversion, ablation, medication initiation or changes) related to AFib following the use of the Apple Watch.

This study will leverage data from large-scale studies like the Heartline Study and the Apple Heart Study to evaluate the Apple Watch's time to diagnosis, and the positive predictive value of irregular rhythm notifications. These studies employ pragmatic, prospective, and often randomized controlled designs, while strengthening statistical power analysis by their large participant size. Inferential statistics can be used to compare diagnostic accuracy metrics (sensitivity, specificity) of the Apple Watch to standard ECG using appropriate statistical tests and confidence intervals. Regression analysis could be used to assess the association between patient characteristics (age, AFib history, technical literacy) and Apple Watch reliability (e.g., number of inconclusive readings) and clinical utilization (e.g., frequency of use, adherence to follow-up). T-tests and chi-square tests can be used to compare healthcare utilization rates to groups of patients using Apple Watch versus those using other methods (Dhruva et al., 2021). A meta-analysis could be conducted to pool estimates of diagnostic accuracy or other outcomes, taking into account potential heterogeneity between studies (Shahid et al., 2025).

Qualitative data will be collected on patient and healthcare provider perceptions of the Apple Watch and its AFib feature. Focus groups and semi-structured interviews will be conducted, including patient forum analysis from online communities and reviews discussing Apple Watch's atrial fibrillation detection. Thematic Analysis would be the most appropriate method to analyze the data and identify common patterns in attitudes, experiences, and concerns regarding Apple Watch's accuracy and usability. Some important themes would be perceived accuracy and trust, impact on the patients' health decisions, limitations of the feature, false

positives, and clinical practice integration. A literature review would be conducted to find existing studies on wearable health technology and atrial fibrillation detection. Surveys or interviews with cardiologists, primary care physicians, and other healthcare professionals can be utilized to gather their perspectives on the utility of Apple Watch data in diagnosing and managing AFib, their concerns about false positives and increased workload, and their recommendations for integrating this technology into clinical practice.

Project Summary: Evaluating the Apple Watch for Atrial Fibrillation Detection

Objective:

Atrial fibrillation (AFib) is a serious cardiac arrhythmia associated with increased risks of stroke, heart failure, and mortality. Early detection of AFib is crucial to preventing complications. Wearable devices like the Apple Watch, equipped with photoplethysmography (PPG) and electrocardiogram (ECG) technology, have emerged as potential tools for detecting irregular heart rhythms. This study aims to evaluate the diagnostic accuracy, reliability, and clinical utility of the Apple Watch in detecting AFib. It will also assess patient and healthcare provider perceptions of the device's efficacy, and its potential impact on healthcare utilization.

Methodology:

- **Quantitative Analysis:**
    - Assessing the sensitivity and specificity of Apple Watch ECG in AFib detection compared to standard 12-lead ECG.
    - Evaluating diagnostic performance metrics from large-scale studies like the Apple Heart Study and Heartline Study.
    - Analyzing healthcare utilization data, including doctor visits, hospitalizations, and diagnostic tests post-AFib detection.
    - Investigating factors affecting accuracy, such as skin tone, tattoos, and physical activity.
- **Qualitative Analysis:**
    - Conducting patient interviews and focus groups to understand user trust, usability concerns, and behavioral changes.

- Surveying healthcare providers to gauge their perceptions of the device's reliability and clinical integration.

Expected Impact:

The findings will contribute to the understanding of consumer-grade wearables in AFib detection, informing future research, regulatory considerations, and clinical adoption. This research will help bridge the gap between technology and healthcare by determining the Apple Watch's role in early AFib detection, and its broader implications for patient care.

References

Canali, S., Schiaffonati, V., & Aliverti, A. (2022). Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. PLOS digital health, 1(10), e0000104. https://doi.org/10.1371/journal.pdig.0000104

Colvonen, P. J., DeYoung, P. N., Bosompra, N. A., & Owens, R. L. (2020). Limiting racial disparities and bias for wearable devices in health science research. Sleep, 43(10), zsaa159. https://doi.org/10.1093/sleep/zsaa159

Dhruva, S. S., Shah, N. D., Vemulapalli, S., Deshmukh, A., Beatty, A. L., Gamble, G. M., Freeman, J. V., Hummel, J. P., Piccini, J. P., Akar, J. G., Ervin, K., Arges, K. L., Emanuel, L., Noseworthy, P. A., Hu, T., Bartlett, V., & Ross, J. S. (2021). Heart Watch Study: protocol for a pragmatic randomised controlled trial. BMJ open, 11(12), e054550. https://doi.org/10.1136/bmjopen-2021-054550

Elbey, M. A., Young, D., Kanuri, S. H., Akella, K., Murtaza, G., Garg, J., Atkins, D., Bommana, S., Sharma, S., Turagam, M., Pillarisetti, J., Park, P., Tummala, R., Shah, A., Koerber, S., Shivamurthy, P., Vasamreddy, C., Gopinathannair, R., & Lakkireddy, D. (2021). Diagnostic Utility of Smartwatch Technology for Atrial Fibrillation Detection - A Systematic Analysis. Journal of atrial fibrillation, 13(6), 20200446. https://doi.org/10.4022/jafib.20200446

Garcia, A., Balasubramanian, V., Lee, J., Gardner, R., Gummidipundi, S., Hung, G., Ferris, T., Cheung, L., Granger, C., Kowey, P., Rumsfeld, J., Russo, A., Hills, M. T., Talati, N., Nag, D., Stein, J., Tsay, D., Desai, S., Mahaffey, K., Turakhia, M., … Desai, M. (2022). Lessons learned in the Apple Heart Study and implications for the data management of future digital clinical trials. Journal of biopharmaceutical statistics, 32(3), 496–510. https://doi.org/10.1080/10543406.2022.2080698

Gibson, C. M., Steinhubl, S., Lakkireddy, D., Turakhia, M. P., Passman, R., Jones, W. S., Bunch, T. J., Curtis, A. B., Peterson, E. D., Ruskin, J., Saxon, L., Tarino, M., Tarakji, K. G., Marrouche, N., Patel, M., Harxhi, A., Kaul, S., Nikolovski, J., Juan, S., Wildenhaus, K., … Heartline Steering Committee (2023). Does early detection of atrial fibrillation reduce the risk of thromboembolic events? Rationale and design of the Heartline study. American heart journal, 259, 30–41. https://doi.org/10.1016/j.ahj.2023.01.004

Heo, R. H., Foroutan, F., De Luca, E., Albertini, L., Rac, V. E., Cafazzo, J. A., Hershman, S. G., Duero Posada, J. G., Ross, H. J., & Moayedi, Y. (2024). Feasibility of Single-Lead Apple Watch Electrocardiogram in Atrial Fibrillation Detection Among Heart Failure Patients. JACC. Advances, 3(7), 101051. https://doi.org/10.1016/j.jacadv.2024.101051

Huhn, S., Axt, M., Gunga, H. C., Maggioni, M. A., Munga, S., Obor, D., Sié, A., Boudo, V., Bunker, A., Sauerborn, R., Bärnighausen, T., & Barteit, S. (2022). The Impact of Wearable Technologies in Health Research: Scoping Review. JMIR mHealth and uHealth, 10(1), e34384. https://doi.org/10.2196/34384

Inocian, E., Junia, A., Cordovez, M. G., Katada, K., & Paday, L. M. W. (2024). Accuracy of the Apple Watch in detecting atrial fibrillation among patients undergoing 24-hour Holter monitoring. European Heart Journal, 45(Supplement_1). https://doi.org/10.1093/eurheartj/ehae666.350

Mohamoud, A., Jensen, J., & Buda, K. G. (2024). Consumer-grade wearable cardiac monitors: What they do well, and what needs work. Cleveland Clinic journal of medicine, 91(1), 23–29. https://doi.org/10.3949/ccjm.91a.23030

Prasitlumkum, N., Cheungpasitporn, W., Chokesuwattanaskul, A., Thangjui, S., Thongprayoon, C., Bathini, T., Vallabhajosyula, S., Kanitsoraphan, C., Leesutipornchai, T., & Chokesuwattanaskul, R. (2021). Diagnostic accuracy of smart gadgets/wearable devices in detecting atrial fibrillation: A systematic review and meta-analysis. Archives of cardiovascular diseases, 114(1), 4–16. https://doi.org/10.1016/j.acvd.2020.05.015

Shahid, S., Iqbal, M., Saeed, H., Hira, S., Batool, A., Khalid, S., & Tahirkheli, N. K. (2025). Diagnostic Accuracy of Apple Watch Electrocardiogram for Atrial Fibrillation: A Systematic Review and Meta-Analysis. JACC. Advances, 4(2), 101538. https://doi.org/10.1016/j.jacadv.2024.101538

Vyas, R., Jain, S., Thakre, A., Thotamgari, S. R., Raina, S., Brar, V., Sengupta, P., & Agrawal, P. (2024). Smart watch applications in atrial fibrillation detection: Current state and future directions. Journal of cardiovascular electrophysiology, 35(12), 2474–2482. https://doi.org/10.1111/jce.16451

Wasserlauf, J., Vogel, K., Whisler, C., Benjamin, E., Helm, R., Steinhaus, D. A., Yousuf, O., & Passman, R. S. (2023). Accuracy of the Apple watch for detection of AF: A multicenter experience. Journal of cardiovascular electrophysiology, 34(5), 1103–1107. https://doi.org/10.1111/jce.15892