Review

Real-World Accuracy of Wearable Activity Trackers for Detecting Medical Conditions: Systematic Review and Meta-Analysis

Ben Singh[1], PhD; Sebastien Chastin[2], PhD; Aaron Miatke[2], MSc; Rachel Curtis[2], PhD; Dorothea Dumuid[2], PhD; Jacinta Brinsley[2], PhD; Ty Ferguson[2], PhD; Kimberley Szeto[2], PhD; Catherine Simpson[2], MSc; Emily Eglitis[2], MSc; Iris Willems[2], MSc; Carol Maher[2], PhD

[1]Allied Health & Human Performance, University of South Australia, Adelaide, Australia

[2]Department of Rehabilitation Sciences and Physiotherapy, Ghent University, Ghent, Belgium

Corresponding Author:
Ben Singh, PhD
Allied Health & Human Performance
University of South Australia
Corner of North Terrace and Frome Road
Adelaide, 5001
Australia
Phone: 61 1300301703
Email: ben.singh@unisa.edu.au

Abstract

Background: Wearable activity trackers, including fitness bands and smartwatches, offer the potential for disease detection by monitoring physiological parameters. However, their accuracy as specific disease diagnostic tools remains uncertain.

Objective: This systematic review and meta-analysis aims to evaluate whether wearable activity trackers can be used to detect disease and medical events.

Methods: Ten electronic databases were searched for studies published from inception to April 1, 2023. Studies were eligible if they used a wearable activity tracker to diagnose or detect a medical condition or event (eg, falls) in free-living conditions in adults. Meta-analyses were performed to assess the overall area under the curve (%), accuracy (%), sensitivity (%), specificity (%), and positive predictive value (%). Subgroup analyses were performed to assess device type (Fitbit, Oura ring, and mixed). The risk of bias was assessed using the Joanna Briggs Institute Critical Appraisal Checklist for Diagnostic Test Accuracy Studies.

Results: A total of 28 studies were included, involving a total of 1,226,801 participants (age range 28.6-78.3). In total, 16 (57%) studies used wearables for diagnosis of COVID-19, 5 (18%) studies for atrial fibrillation, 3 (11%) studies for arrhythmia or abnormal pulse, 3 (11%) studies for falls, and 1 (4%) study for viral symptoms. The devices used were Fitbit (n=6), Apple watch (n=6), Oura ring (n=3), a combination of devices (n=7), Empatica E4 (n=1), Dynaport MoveMonitor (n=2), Samsung Galaxy Watch (n=1), and other or not specified (n=2). For COVID-19 detection, meta-analyses showed a pooled area under the curve of 80.2% (95% CI 71.0%-89.3%), an accuracy of 87.5% (95% CI 81.6%-93.5%), a sensitivity of 79.5% (95% CI 67.7%-91.3%), and specificity of 76.8% (95% CI 69.4%-84.1%). For atrial fibrillation detection, pooled positive predictive value was 87.4% (95% CI 75.7%-99.1%), sensitivity was 94.2% (95% CI 88.7%-99.7%), and specificity was 95.3% (95% CI 91.8%-98.8%). For fall detection, pooled sensitivity was 81.9% (95% CI 75.1%-88.1%) and specificity was 62.5% (95% CI 14.4%-100%).

Conclusions: Wearable activity trackers show promise in disease detection, with notable accuracy in identifying atrial fibrillation and COVID-19. While these findings are encouraging, further research and improvements are required to enhance their diagnostic precision and applicability.

Trial Registration: Prospero CRD42023407867; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=407867

KEYWORDS

wearable activity trackers; disease detection; atrial fibrillation; COVID-19 diagnosis; meta-analysis; wearables; wearable tracker; tracker; detection; monitoring; physiological; diagnostic tool; tool; tools; Fitbit; atrial; COVID-19; wearable

Introduction

As health care budgets around the world continue to soar, the need for cost-effective interventions that both reduce health care costs and improve patient outcomes has never been more urgent [1]. Early detection of medical conditions offers a pathway to achieve these goals, enabling prompt intervention during acute medical events or even pre-emptive action before such events occur [2]. Wearable activity monitors are emerging as a potential tool in this evolving landscape.

In recent years, wearable activity trackers have become ubiquitous tools, widely adopted for tracking and enhancing physical activity and other lifestyle behaviors, helping to mitigate the risk of chronic diseases [3]. These devices measure a plethora of activity metrics such as steps taken, distance covered, energy expenditure, physical activity intensities, and sleep patterns [4]. The scientific literature has witnessed a surge in original studies and systematic reviews and meta-analyses, focused on determining the reliability and validity of activity trackers for measuring activity levels [5,6] and their effectiveness in intervening in daily activity patterns and downstream health outcomes [7-12]. These studies have shown that interventions using consumer-based wearable activity trackers can increase physical activity participation and lead to significant improvements in health outcomes, across a range of populations [7-12]. As wearable technology has progressed, wearable activity trackers offer increasing potential to move beyond activity metrics and aid in the early identification of diseases and other medical events.

Rapid technological advancements have significantly extended the capabilities of contemporary consumer-grade wearable activity trackers such as Fitbits and Apple Watches [13]. Modern wearables incorporate sophisticated sensors capable of monitoring a wide array of physiological parameters beyond just movement including heart rate, blood oxygen levels, sleep quality, and stress markers [14]. While this expanded functionality holds promise for disease detection and monitoring, the evidence supporting the use of consumer wearables for such applications remains limited. For example, the systematic review by Alban-Cadena et al [15] evaluated wearable sensors for monitoring Parkinson disease–related gait impairments and symptoms such as tremors, bradykinesia, and dyskinesia. However, most included studies were very small (10-20 participants) and were conducted in controlled laboratory environments using specialized setups such as multi-sensor accelerometer arrays worn on the ankles and spine. While offering the potential for home-based rehabilitation, the generalizability of these findings to widely adopted, consumer-oriented wearable trackers designed for real-world, free-living conditions is unclear.

Other recent systematic reviews have evaluated the accuracy of wearable tracking devices for detecting specific health conditions such as arrhythmias [16], cardiovascular disease [17,18], and COVID-19 [19]. However, these reviews have notable limitations. Most included studies were conducted in controlled laboratory settings, limiting the generalizability of their findings to real-world, free-living conditions [16,17,19].

Additionally, these reviews focused narrowly on individual clinical outcomes, preventing comparisons of wearables' detection accuracy across different medical conditions and events. For example, the narrative syntheses highlighted wearables' potential as complementary tools for detecting cardiovascular conditions such as arrhythmias, atrial fibrillation, myocardial infarction, and heart failure [16,17]. The meta-analysis of Lee et al [18] of 26 studies found wearable devices had a pooled sensitivity of 94.80% and specificity of 96.96% for atrial fibrillation detection. In contrast, Cheong et al [19] reported lower diagnostic accuracy for COVID-19 detection, with area under the curve (AUC) values ranging from 75% to 94.4% and sensitivity and specificity ranging from 36.5% to 100% and 73% to 95.3%, respectively. Notably, all but one review [18] used narrative synthesis approaches [16,17,19], limiting their ability to quantify detection accuracy, and preventing readers from comparing detection accuracy across conditions reported in the respective reviews.

As wearable technology rapidly evolves, with frequent introductions of new and more advanced devices, the scientific evidence base for disease detection is growing, encompassing a wider range of medical conditions and events. Consequently, there is now sufficient data to warrant a comprehensive systematic review with meta-analyses, allowing quantitative comparisons of wearables' detection accuracy across various conditions in real-world settings.

Our systematic review and meta-analysis aim to fill this crucial gap by comprehensively assessing the reliability and accuracy of consumer-grade wearable activity trackers for detecting and monitoring a wide range of medical conditions and events in free-living, real-world settings. Unlike previous reviews that relied on narrative synthesis approaches, our quantitative meta-analyses will allow for robust comparisons of wearables' diagnostic performance across diverse conditions and events. By rigorously evaluating evidence from studies conducted in real-world contexts, our review will provide evidence to guide the responsible and effective implementation of wearable technology for early detection and continuous health monitoring by researchers, health care providers, policy makers, technology companies, and other stakeholders. As consumer adoption of wearables continues to rise rapidly worldwide, our comprehensive synthesis will assist in harnessing their potential while mitigating risks and ensuring appropriate use.

Methods

Protocol and Registration

The protocol for this systematic review was prospectively registered on PROSPERO (ID CRD42023407867) and this paper is reported according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [20] guidelines.
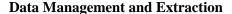
Selection Criteria and Search Strategy

The inclusion criteria are summarized in Table S1 in Multimedia Appendix 1. The inclusion criteria were developed using the population, exposure, outcomes, and study type criteria as follows: adult population (aged 18 years or older) in free-living

conditions, that have not been recruited based on a specific health condition or diagnosis; use of a wearable activity tracker (eg, Fitbit, Apple Watch, or a research-grade accelerometer) for the detection of any disease or medical event (eg, atrial fibrillation, the onset of infectious disease, and falls). To be eligible, the wearable activity tracker had to be able to detect movement behavior (ie, include an accelerometer), but could also include other types of sensors (eg, light sensor and temperature sensor). The wearable activity tracker had to consist of a single device worn on a single body location (eg, on the wrist or chest, not across both); studies needed to assess the actual diagnosis of a medical condition or occurrence of events that had clinical relevance (eg, falls). Eligible studies are needed to report an outcome related to diagnostic accuracy, such as specificity or sensitivity of the device for early detection of disease or medical events. Examples could include but were not limited to, providing effect estimates of overall diagnostic accuracy (%), sensitivity (%), and specificity (%) with 95% CIs; and validation studies conducted under free-living conditions that were reported in a peer-reviewed journal study were included. This included secondary analyses conducted within the context of observational studies, experimental studies, or quasi-experimental studies. Both consumer-initiated studies, where existing consumers who had purchased their own wearables were invited to join a study, and researcher-initiated studies, where researchers recruited participants and provided them with wearables, were included, as they represent 2 complementary real-world contexts in which wearable devices are often implemented for disease detection and monitoring. Studies were included only if they evaluated wearable devices provided by health care providers or researchers as part of a formal monitoring program, and the detection of a specific clinical event or disease was a prespecified outcome measure of the study. Studies examining consumer-driven self-tracking with personal wearables outside of a health care context were excluded. The following were also excluded: studies involving children or adolescents, studies examining symptoms within people known to have a specific disease, wearable devices that cannot track activity levels (eg, continuous glucose monitors), studies evaluating an array of wearable sensors worn at multiple body locations (eg, watch plus skin patch) or pedometers, studies measuring the association between an exposure and an outcome (eg, using odds ratios, relative risk, and hazard ratios), lab- or hospital-based studies, and conference abstracts or dissertations.

Ten databases were searched (CINAHL, Cochrane Library, Embase via OVID, MEDLINE via OVID, Emcare via OVID, JMIR Publications, ProQuest Central, ProQuest Nursing and Allied Health Source, PsycINFO, and Scopus) using subject heading, keyword, and MeSH (Medical Subject Headings) term searches for terms related to "wearable device" and "detection" (see Table S2 in Multimedia Appendix 1 for the full search strategy). We intentionally used broad search terms to ensure a comprehensive capture of the evidence base, including all types of medical conditions and events, without restricting our search to predefined diagnostic or event outcomes. Database searches were limited to peer-reviewed journal studies published in English from inception to April 1, 2023.

Data Management and Extraction

Search results were imported into ASReview (version 2.0; ASReview Community), an open-source software artificial intelligence tool designed for screening studies for systematic reviews. Title or abstract screening was conducted in ASReview by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE). The software uses an active learning algorithm that iteratively selects the most relevant studies for inclusion based on the initial judgments made by the research team. The screening was stopped when 100 consecutive nonrelevant studies were screened. Following title or abstract screening, results were then imported to EndNote X9 (Clarivate) where duplicates were removed and then exported into Covidence (Veritas Health Innovation) for full-text screening, data extraction, and risk of bias scoring which was completed in duplicate by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE), with disagreements resolved by discussion.

Data were extracted in duplicate by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE) using a standardized extraction form in Covidence. The risk of bias in the included reviews was assessed by 2 independent reviewers in duplicate using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Diagnostic Test Accuracy Studies. Studies were rated out of nine for the following items: (1) enrollment of consecutive or random sample, (2) the avoidance of a case-control design, (3) inappropriate exclusions, (4) the interpretation of index test results, (5) the prespecification of thresholds, (6) reference standard classification, (7) the interpretation of reference standard, (8) timing of tests, and (9) analysis.

Data Synthesis and Analysis

For each meta-analysis, data were combined at the study level. Separate meta-analyses were performed for (1) COVID-19 detection, (2) atrial fibrillation or arrhythmia detection, and (3) fall detection. Outcomes of interest were analyzed and data were pooled using sensitivity (%), specificity (%), AUC (%), accuracy (%), and positive predictive value (PPV), with 95% CIs as the effects measures. Sensitivity (%) denotes the percentage of individuals with the disease or condition correctly identified by the test, while specificity (%) represents the percentage of those without the disease or condition correctly identified as negative. The AUC (%) quantifies the test's overall diagnostic accuracy, ranging from 0% to 100%, with higher values indicating better performance. Accuracy (%) reflects the proportion of all tests accurately classified, and PPV (%) indicates the likelihood that a positive test result correlates with the disease or condition being tested for. If 95% CIs were not reported in a study, they were calculated based on available data, using recommended formulas [21]. Publication bias was evaluated using funnel plots of effect sizes and standard errors and evaluating for asymmetries or missing sections within the plot, for meta-analyses that involved more than 10 studies. The Cochran Q test was used to assess statistical heterogeneity and the I^2 statistic was used to quantify the proportion of the overall outcome attributed to variability. The following cut-off values for the I^2 statistic were used: 0% to 29%=no heterogeneity; 30%

to 49%=moderate heterogeneity; 50% to 74%=substantial heterogeneity; and 75% to 100%=considerable heterogeneity [22]. Subgroup analyses were undertaken to evaluate device type (Fitbit, Apple watch, Oura ring, and others) for outcomes that had at least 2 studies in each subgroup. Sensitivity analyses for the meta-analysis were performed by removing the study with the lowest sensitivity, specificity, AUC, accuracy, or PPV. All meta-analyses were performed using Stata/MP (version 16; StataCorp).

The overall level of evidence was graded using the Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence, as follows: grade A: consistent level 1 studies (ie, individual randomized controlled trials); B: consistent level 2 (ie, individual cohort studies) or 3 studies (ie, individual case-control studies) or extrapolations from level 1 studies; C: level 4 studies (ie, case series) or extrapolations from level 2 or 3 studies; or D: level 5 (ie, expert opinion without explicit critical appraisal) evidence or inconsistent or inconclusive studies of any level [23]. Each outcome of interest was assigned a "Grade of Recommendation" based on meeting these criteria.

Deviations From the Registered Protocol

We planned to use the Effective Public Health Practice Project Quality Assessment Tool to assess study quality and risk of bias. However, during data extraction and quality assessment, we opted to use the JBI Critical Appraisal Checklist for Diagnostic Test for Accuracy Studies, as this instrument was more relevant to the included studies. Further, we were unable to conduct subgroup analyses for the type of wearable for atrial fibrillation and fall detection, due to an insufficient number of studies.
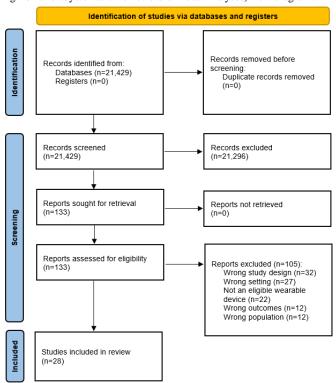
Results

Overview

Of the 21,429 records identified following the database search, 28 were eligible (see Figure 1 for PRISMA flowchart including reasons for exclusions; see Table S3 in Multimedia Appendix 1 for a complete list of full texts that were excluded during the final stage of screening, with reasons). An overview of all included study's characteristics is shown in Table S4 in Multimedia Appendix 1. There was a total of 1,226,801 participants (median sample size 264, IQR 96-8338; range 29-455,699). Median participant age was 47.3 (IQR 36.6-66), between 28.6 and 78.3, years and 21 (75%) studies involved female and male participants (gender was not reported in 7 (25%) studies). A total of 16 (57%) studies evaluated COVID-19, 5 (18%) studies evaluated atrial fibrillation, 3 (11%) studies assessed a broad range of cardiac arrhythmias, 3 (11%) studies assessed falls, and 1 (3.6%) study assessed viral symptoms. The devices used in the studies were Fitbit (n=6), Apple Watch (n=6), Oura ring (n=3), a combination of various devices (ie, studies that used a combination of the Apple Watch, Fitbit, Garmin, and other devices; n=7), Empatica E4 (n=1), Dynaport MoveMonitor (n=2), Samsung Galaxy Watch (n=1), and other or not specified (n=2). The median score for the JBI Critical Appraisal Checklist for Diagnostic Test Accuracy Studies was 6 (IQR 5-7; range 1-9) out of 9 (Table S5 in Multimedia Appendix 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

There was sufficient data in the included studies to conduct meta-analyses for the following clinimetrics: (1) COVID-19 detection (accuracy, %; sensitivity, %; AUC, %; and specificity, %), (2) atrial fibrillation detection (PPV, %; sensitivity, %; and specificity, %), and (3) falls detection (sensitivity, %; and specificity, %).

Meta-Analysis Results

COVID-19 Detection

Meta-analysis results of AUC, accuracy, sensitivity, and specificity for COVID-19 detection are shown in Figure 2. Meta-analyses of 9 studies showed a pooled AUC of 80.15% (95% CI 71.03%-89.27%) and 5 studies had a pooled accuracy of 87.54% (95% CI 81.57%-93.51%). Pooled sensitivity from 8 studies was 79.53% (95% CI 67.73%-91.33%), and 7 studies showed a pooled specificity of 76.79% (95% CI 69.44%-84.13%).

Subgroup analysis for device type for sensitivity and specificity are shown in Figures S6 and S7 in Multimedia Appendix 1,

respectively. A summary of sensitivity and specificity for the different devices is shown in Figure 3. Overall, the Fitbit had a sensitivity and specificity of 75.39% and 90.60%, respectively, the Oura ring had a sensitivity and specificity of 80.47% and 72.60%, respectively, and combined devices had a sensitivity and specificity of 82.69% and 74.62%, respectively.

The results of sensitivity analyses are shown in Figure S3 in Multimedia Appendix 1. Following the removal of the worst-performing study, AUC was 84.10%, accuracy was 88.65%, sensitivity was 85.62%, and specificity was 78.57%.

Grade of recommendation: (B) consistent level 2 studies supporting the use of wearable activity trackers for the detection of COVID-19.

Figure 2. Meta-analysis of accuracy, sensitivity, AUC, and specificity of wearable activity trackers for detection of COVID-19. AUC: area under the curve.
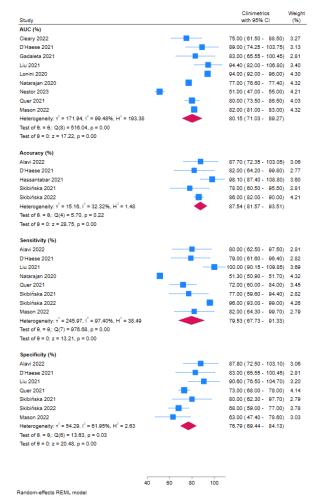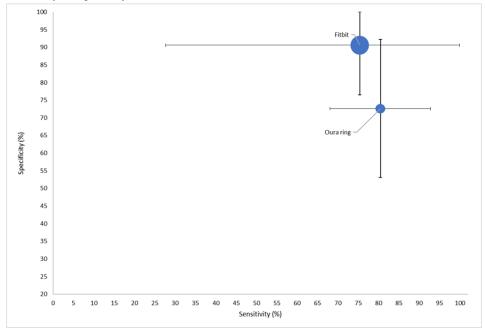
Figure 3. Overview of sensitivity and specificity for the different devices for COVID-19 detection.
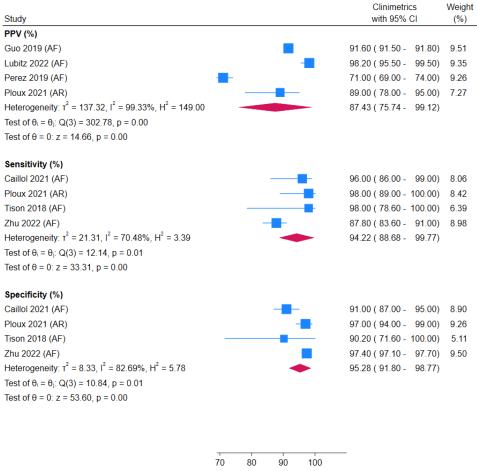

Atrial Fibrillation Detection

Pooled analyses of PPV, sensitivity, and specificity for atrial fibrillation detection are shown in Figure 4. Meta-analysis of 4 studies showed a combined PPV of 87.43% (95% CI 75.74%-99.12%). Pooled sensitivity was 94.22% (95% CI 88.68%-99.77%; 4 studies) and pooled specificity was 95.28% (95% CI 91.80%-98.77%; 4 studies).

The results of sensitivity analyses are shown in Figure S4 in Multimedia Appendix 1. Following the removal of the worst-performing study, PPV was 93.64%, sensitivity was 97.28%, and specificity was 95.55%.

Grade of recommendation: (B) consistent level 2 studies supporting the use of wearable activity trackers for the detection of atrial fibrillation.

Figure 4. Meta-analysis of PPV, sensitivity, and specificity of wearable activity trackers for detection of AF and AR. AF: atrial fibrillation; AR: arrhythmia; PPV: positive predictive value.
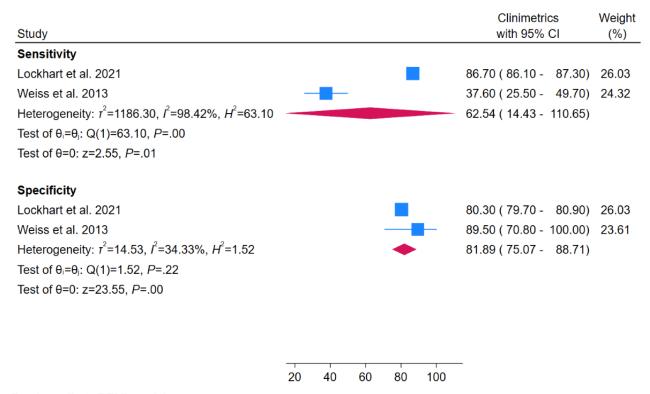

Falls Detection

Meta-analysis results of sensitivity and specificity for fall detection are shown in Figure 5. Meta-analyses of 2 studies showed a specificity of 62.54% (95% CI 14.43%-100%) and a sensitivity of 81.89% (95% CI 75.07%-88.17%). There was an insufficient number of studies for subgroup analyses of device type and sensitivity analyses for fall detection.

Grade of recommendation: (D) inconsistent or inconclusive studies of any level for the use of wearable activity trackers to predict falls.

Figure 5. Meta-analyses of sensitivity and specificity of wearable activity trackers for detection for fall detection.

Discussion

Principal Findings

In this study, we set out to systematically review and meta-analyze the current evidence regarding wearable activity trackers' ability to detect medical conditions and events under free-living conditions. To date, the majority of studies have focused on the detection of COVID-19, with a smaller number of studies focused on cardiac conditions and falls. For COVID-19 detection, the devices generally demonstrated good sensitivity and specificity. The most promising results were found for the detection of atrial fibrillation, for which the wearables showed high sensitivity and specificity. Whereas, for fall detection, the present findings devices showed moderate sensitivity but lower specificity. These findings indicate that while these devices are becoming more dependable for monitoring specific health conditions, their performance varies depending on the condition being detected.

The current body of evidence on the diagnostic potential of wearable activity trackers is notably skewed toward COVID-19 detection, a focus that is understandable given the pandemic's global impact and the consequent urgent need for monitoring solutions. Researching the feasibility of detecting COVID-19 through wearables holds appeal due to the availability of widely used reference standards. Rapid and polymerase chain reaction tests, widely used, allow for easy self-reporting of COVID-19 diagnoses by many individuals. In contrast, accessing a reliable gold standard for other health outcomes poses significant challenges. However, what was surprising to note is the limited number of studies exploring these trackers for other health conditions, especially given that numerous wearables advertise

features such as sleep apnea detection—a topic noticeably absent in our findings. Our extensive database search identified only a handful of studies each related to cardiac issues and falls. This gap in the literature is striking considering the wide array of health conditions that could theoretically be monitored using wearable technology, given their ability to capture data related to heart rate, movement, skin temperature, and more. Such capabilities would suggest that a broad spectrum of medical conditions could be measured, spanning cardiovascular and respiratory conditions to neurological and psychological disorders. It is important to note that we intentionally focused on the accuracy of data collected in free-living conditions (with a view to understanding current-day diagnostic capabilities). We note numerous laboratory-based studies that were excluded (eg, [24] and [25]) suggesting that a wider range of diagnostic outcomes may become available in the future. Furthermore, many studies were excluded because they focused on monitoring symptoms in people with a known diagnosis (eg, seizures in people with epilepsy [26], and freezing gait in Parkinson disease [27]) which was outside the scope of this study, but highlights wearable activity trackers' potential for medical condition monitoring.

This study revealed that wearable activity trackers demonstrate moderate to high sensitivity and specificity for COVID-19 detection. It is interesting to compare our results with those for other COVID-19 screening tests. A systematic review by Mistry et al [28] on lateral flow devices (LFD) tests (also known as rapid antigen tests) evaluated 24 papers across 8 different LFD brands, covering over 26,000 test results. Their findings indicated that sensitivity ranged from 37.7% to 99.2% and specificity ranged from 92.4% to 100% [28]. Comparatively,

this study's pooled sensitivity for wearable-detected COVID-19 was 79.5% (range 51.3%-100%), which is in line with the LFD results. However, our specificity of 76.8% (range 63%-90.6%) was slightly lower. According to UK government guidelines, the benchmarks for COVID-19 workplace screening are ≥68% for sensitivity and ≥97% for specificity [29]. This suggests that while wearable activity monitor detection meets the sensitivity criterion, it falls short on specificity.

The most promising results were observed for the detection of atrial fibrillation, with figures that compare favorably to other clinical tests. For example, the sensitivity and specificity of a 12-lead electrocardiogram for detecting atrial fibrillation have previously been shown to range between 93% and 97% [30,31], which appears similar to our sensitivity and specificity of 94.2% and 95.3%, respectively. Over the course of 2022-2023, major brands, such as Fitbit [32], Apple Watch [33], Garmin [34], and Samsung [35], received approval from the US Food and Drug Administration for their atrial fibrillation detection features. The relatively higher accuracy in identifying cardiac arrhythmias as compared to COVID-19 is perhaps expected, given that cardiac functions can be deduced from wearables' optical heart-rate sensors. In contrast, COVID-19 detection usually requires intricate algorithms that amalgamate multiple data points [36,37].

While wearable activity trackers demonstrated effectiveness in detecting cardiac arrhythmia and COVID-19, our meta-analysis revealed that their accuracy in detecting falls was only moderate. The devices were generally effective in identifying actual falls, with a sensitivity of 81.9%. However, they also generated a significant number of false positives, as evidenced by a lower specificity of 62.5%. This aligns with existing literature on the subject [38,39]. It is crucial to note that our review specifically focused on the performance of these devices in real-world conditions among the general population. Most existing studies on fall detection with wearables have been conducted in controlled laboratory settings using simulated falls, where accuracy has generally been higher [38,39]. The false positives in fall detection are likely due to the devices relying on accelerometry data, which can misinterpret other rapid downward movements as falls. Further research is needed to refine the algorithms used in these devices to improve their performance in fall detection. Future studies might incorporate additional metrics, such as rapid changes in heart rate or galvanic skin response, which may accompany a fall, to enhance accuracy.

This study offers several significant strengths, including being the first systematic review and meta-analysis focused on the real-world accuracy of wearable activity trackers in detecting medical conditions and events. The review analyzed a robust data set from 28 studies, involving over 1 million participants, enabling a comprehensive meta-analysis of various outcomes. Instead of limiting our focus to specific diagnostic outcomes, we examined a broad range of medical conditions. Our search strategy was exceptionally thorough, encompassing 10 databases and reviewing over 21,000 studies to capture a wide array of diagnostic outcomes. Methodologically, we adhered to the PRISMA 2020 guidelines, which included conducting sensitivity and subgroup analyses, as well as evaluating the certainty of the evidence.

Study limitations must be acknowledged. There was considerable heterogeneity in the designs of included studies, such as their reference standards, diagnostic tests, and sample characteristics. Given the size of the current evidence, there were too few studies to conduct separate subgroup analyses based on specific device models or software versions. Our review included both researcher-initiated and consumer-initiated studies to provide a comprehensive assessment of wearable activity trackers in real-world settings. Researcher-initiated studies typically involved smaller sample sizes and controlled participant recruitment, while consumer-initiated studies often had larger sample sizes and reflected more naturalistic use patterns. While this combination enhances the generalizability of our findings, it also introduces heterogeneity. We acknowledge this as a limitation and suggest that future research should consider these differences when interpreting results. Additionally, our review only identified studies in the domains of COVID-19, cardiovascular conditions, and falls as eligible. While laboratory-based studies are being conducted for event detection in other health domains (such as stress and respiratory conditions), our focus was intentionally on studies conducted in free-living conditions. This approach offers insights into the wearables' event detection capabilities in real-world settings, as opposed to artificial (eg, laboratory) conditions.

Clinical Implications

The use of wearable activity trackers for detecting medical events is an emerging field with both significant promise and challenges. Wearable activity trackers demonstrate comparable ability to detect COVID-19 and atrial fibrillation compared with other clinical tests such as lateral flow tests and electrocardiograms. However, wearables offer the additional advantage of continuous, real-time monitoring for conditions requiring constant surveillance. As such, they may empower patients to take a more proactive role in their health care by giving them immediate feedback and data about their condition. They may also contribute to improved surveillance and resource planning for health care systems, which could be particularly useful in times of epidemics or pandemics.

Certain wearable device features excel at detecting specific medical events. For COVID-19, devices combining heart rate monitors, skin temperature sensors, and accelerometers proved effective by detecting deviations from an individual's baseline across multiple physiological parameters. In contrast, for atrial fibrillation detection, Food and Drug Administration-approved devices relied on optical heart rate sensors providing photoplethysmography data, capable of identifying irregular heart rhythms characteristic of arrhythmias. Fall detection primarily uses accelerometer data, with wrist-worn placement crucial for sensing sudden deceleration and impact forces. However, false positives persist due to nonfall rapid movements. Looking ahead, integrating multiple sensors can enhance accuracy across various medical conditions. Yet, fundamental sensor limitations may remain. Aligning device capabilities with specific use cases and recognizing sensor shortcomings

will inform future research and benchmarking efforts amid evolving technology.

As consumer wearables gradually morph from being lifestyle tools to over-the-counter medical instruments, they present a range of challenges, including concerns about data privacy and security, which will require stringent protective measures. Furthermore, as wearable devices become increasingly sophisticated in detecting medical conditions, such as atrial fibrillation, they offer both benefits and pitfalls. On the positive side, these devices have the potential to identify asymptomatic atrial fibrillation episodes. This is enormously beneficial, since currently, stroke is the first manifestation in at least 25% of atrial fibrillation-related stroke cases [40]. Early detection could therefore lead to timely intervention and stroke prevention. However, health care professionals have reported an uptick in patient consultations triggered by atrial fibrillation alerts from wearables, resulting in a surge of medical tests, such as electrocardiograms, to confirm diagnoses [41]. While some clinicians see this as an advancement in patient-initiated health care, others question the necessity of such screening, particularly in patient subgroups where atrial fibrillation may have a relatively benign prognosis [42]. Moreover, the use of wearables can generate both false positives and negatives, potentially causing unnecessary anxiety, diagnostic tests, and treatments, or giving users a false sense of security.

Future Research

Our review reveals that the current peer-reviewed evidence base concerning the event detection capabilities of consumer wearable activity trackers in free-living conditions is limited to COVID-19, cardiac function, and falls. This was somewhat surprising, given the potential of these devices to diagnose numerous other conditions. Our findings indicate a significant gap in the current literature, which was not apparent in previous reviews that typically focused on specific conditions and did not highlight the lack of studies across a broader range of conditions. Considering the diverse array of sensors incorporated in modern wearable activity trackers, these devices offer considerable potential for detecting and monitoring medical events across an extensive spectrum of health conditions into the future. This may include respiratory conditions, neurological disorders, mental health, stress and fatigue, and even environmental and allergic reactions. This will require research across the product design continuum, from algorithm training to laboratory testing and free-living testing. This will be made all the more challenging by the rapid pace at which new devices and models are released into the market. In the future, our meta-analysis could be updated to provide insight into the accuracy of such diagnostics by condition, device, and population.

Conclusions

This study provides a comprehensive overview of the current state of evidence regarding the diagnostic capabilities of consumer wearable activity trackers in real-world settings. While the devices show promise in detecting conditions, such as COVID-19 and atrial fibrillation, with moderate to high sensitivity and specificity, their performance in detecting falls is moderate, highlighting the need for further refinement of detection algorithms. The existing literature is notably skewed toward COVID-19, leaving a significant gap in our understanding of how these devices can be used for a broader range of health issues. This gap, which was not apparent in previous reviews, underscores the necessity for future research to expand the scope of conditions studied. As wearable technology continues to evolve, it is crucial to address the challenges posed by false positives and negatives, data privacy, and security concerns. This will ensure that the rapid advancements in this field can be matched by robust scientific validation, enabling these devices to realize their full potential as tools for health care monitoring and intervention.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 1.

Authors' Contributions

All authors contributed to the review protocol. BS, SC, AM, RC, DD, JB, TF, KS, CS, EE, IW, and CM designed the search strategy and selected studies. BS, AM, RC, DD, JB, TF, KS, CS, EE, and IW extracted the data. BS analyzed the data. BS, SC, and CM drafted the manuscript. All authors contributed to the drafting of the review. All authors revised the manuscript critically for important intellectual content. All authors approved the final version of the article. All authors had access to all the data in the study and could take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.

XSL•FO

RenderX

[DOCX File , 123 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) checklist.
[PDF File (Adobe PDF File), 46 KB-Multimedia Appendix 2]

References

1. Watson SI, Sahota H, Taylor CA, Chen YF, Lilford RJ. Cost-effectiveness of health care service delivery interventions in low and middle income countries: a systematic review. Glob Health Res Policy. 2018;3:17. [FREE Full text] [doi: 10.1186/s41256-018-0073-z] [Medline: 29930989]
2. Wang W, Yan Y, Guo Z, Hou H, Garcia M, Tan X, et al. All around suboptimal health—a joint position paper of the suboptimal health study consortium and European association for predictive, preventive and personalised medicine. EPMA J. 2021;12(4):403-433. [FREE Full text] [doi: 10.1007/s13167-021-00253-2] [Medline: 34539937]
3. Natalucci V, Marmondi F, Biraghi M, Bonato M. The effectiveness of wearable devices in non-communicable diseases to manage physical activity and nutrition: where we are? Nutrients. 2023;15(4):913. [FREE Full text] [doi: 10.3390/nu15040913] [Medline: 36839271]
4. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform. 2019;93:103153. [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
5. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil. 2015;7:24. [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
6. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. 2015;12:159. [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
7. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health. 2022;4(8):e615-e626. [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
8. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth. 2020;8(7):e15576. [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]
9. Szeto K, Arnold J, Singh B, Gower B, Simpson CEM, Maher C. Interventions using wearable activity trackers to improve patient physical activity and other outcomes in adults who are hospitalized: a systematic review and meta-analysis. JAMA Netw Open. 2023;6(6):e2318478. [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.18478] [Medline: 37318806]
10. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med. 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
11. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci. 2022;11(2):184-193. [FREE Full text] [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]
12. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of wearable activity trackers to improve physical activity behavior in patients with rheumatic and musculoskeletal diseases: a systematic review and meta-analysis. Arthritis Care Res (Hoboken). 2019;71(6):758-767. [doi: 10.1002/acr.23752] [Medline: 30221489]
13. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol. 2021;18(8):581-599. [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
14. Shei RJ, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? Eur J Appl Physiol. 2022;122(9):1975-1990. [FREE Full text] [doi: 10.1007/s00421-022-04951-1] [Medline: 35445837]
15. Albán-Cadena AC, Villalba-Meneses F, Pila-Varela KO, Moreno-Calvo A, Villalba-Meneses CP, Almeida-Galárraga DA. Wearable sensors in the diagnosis and study of Parkinson's disease symptoms: a systematic review. J Med Eng Technol. 2021;45(7):532-545. [doi: 10.1080/03091902.2021.1922528] [Medline: 34060967]
16. Pay L, Yumurtaş A, Satti DI, Hui JMH, Chan JSK, Mahalwar G, et al. Arrhythmias beyond atrial fibrillation detection using smartwatches: a systematic review. Anatol J Cardiol. 2023;27(3):126-131. [FREE Full text] [doi: 10.14744/AnatolJCardiol.2023.2799] [Medline: 36856589]
17. Huang JD, Wang J, Ramsey E, Leavey G, Chico TJA, Condell J. Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review. Sensors (Basel). 2022;22(20):8002. [FREE Full text] [doi: 10.3390/s22208002] [Medline: 36298352]

18. Lee S, Chu Y, Ryu J, Park YJ, Yang S, Koh SB. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. Yonsei Med J. 2022;63(Suppl):S93-S107. [FREE Full text] [doi: 10.3349/ymj.2022.63.S93] [Medline: 35040610]

19. Cheong SHR, Ng YJX, Lau Y, Lau ST. Wearable technology for early detection of COVID-19: a systematic scoping review. Prev Med. 2022;162:107170. [FREE Full text] [doi: 10.1016/j.ypmed.2022.107170] [Medline: 35878707]

20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]

21. Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. Comput Biol Med. 2000;30(3):127-134. [doi: 10.1016/s0010-4825(00)00006-8] [Medline: 10758228]

22. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. Cochrane. 2023. URL: http://www.training.cochrane.org/handbook [accessed 2024-08-06]

23. Oxford Centre for Evidence-Based Medicine: levels of evidence (March 2009). University of Oxford; 2022. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 [accessed 2024-08-06]

24. Johansson D, Ohlsson F, Krýsl D, Rydenhag B, Czarnecki M, Gustafsson N, et al. Tonic-clonic seizure detection using accelerometry-based wearable sensors: a prospective, video-EEG controlled study. Seizure. 2019;65:48-54. [FREE Full text] [doi: 10.1016/j.seizure.2018.12.024] [Medline: 30611010]

25. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. Sleep. 2020;43(7):zsaa045. [FREE Full text] [doi: 10.1093/sleep/zsaa045] [Medline: 32215550]

26. Kusmakar S, Karmakar CK, Yan B, Obrien T, Muthuganapathy R, Palaniswami M. Automated detection of convulsive seizures using a wearable accelerometer device. IEEE Trans Biomed Eng. 2019;66(2):421-432. [doi: 10.1109/TBME.2018.2845865] [Medline: 29993501]

27. Ahlrichs C, Samà A, Lawo M, Cabestany J, Rodríguez-Martín D, Pérez-López C, et al. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. Med Biol Eng Comput. 2016;54(1):223-233. [doi: 10.1007/s11517-015-1395-3] [Medline: 26429349]

28. Mistry DA, Wang JY, Moeser ME, Starkey T, Lee LYW. A systematic review of the sensitivity and specificity of lateral flow devices in the detection of SARS-CoV-2. BMC Infect Dis. 2021;21(1):828. [FREE Full text] [doi: 10.1186/s12879-021-06528-3] [Medline: 34407759]

29. UK Government. Technologies validation group: using tests to detect COVID-19. 2021. URL: https://www.gov.uk/guidance/technologies-validation-group-using-tests-to-detect-covid-19 [accessed 2021-10-19]

30. Welton NJ, McAleenan A, Thom HH, Davies P, Hollingworth W, Higgins JP, et al. Screening strategies for atrial fibrillation: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2017;21(29):1-236. [FREE Full text] [doi: 10.3310/hta21290] [Medline: 28629510]

31. Kvist LM, Vinter N, Urbonaviciene G, Lindholt JS, Diederichsen ACP, Frost L. Diagnostic accuracies of screening for atrial fibrillation by cardiac nurses versus radiographers. Open Heart. 2019;6(1):e000942. [FREE Full text] [doi: 10.1136/openhrt-2018-000942] [Medline: 30997131]

32. Google. New Fitbit feature makes AFib detection more accessible. Google Blog. 2022. URL: https://blog.google/products/fitbit/irregular-heart-rhythm-notifications/ [accessed 2023-10-03]

33. Kritz F. Apple Watches have two new FDA-cleared health applications. 2022. URL: https://array.aami.org/content/news/apple-watches-have-two-new-fda-cleared-health-applications [accessed 2023-10-03]

34. Malik A. Garmin launches a new FDA-cleared ECG app for the Venu 2 Plus. 2023. URL: https://tinyurl.com/57y86ccf [accessed 2023-10-03]

35. Samsung announces FDA-cleared irregular heart rhythm notification for Galaxy Watch. Samsung Newsroom US. 2023. URL: https://news.samsung.com/us/fda-cleared-irregular-heart-rhythm-notification-for-galaxy-watch/2023 [accessed 2023-10-03]

36. Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med. 2020;3(1):156. [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]

37. Hassantabar S, Stefano N, Ghanakota V, Ferrari A, Nicola GN, Bruno R, et al. CovidDeep: SARS-CoV-2/COVID-19 test based on wearable medical sensors and efficient neural networks. IEEE Trans Consumer Electron. 2021;67(4):244-256. [doi: 10.1109/tce.2021.3130228]

38. Chen M, Wang H, Yu L, Yeung EHK, Luo J, Tsui K, et al. A systematic review of wearable sensor-based technologies for fall risk assessment in older adults. Sensors (Basel). 2022;22(18):6752. [FREE Full text] [doi: 10.3390/s22186752] [Medline: 36146103]

39. Warrington DJ, Shortis EJ, Whittaker PJ. Are wearable devices effective for preventing and detecting falls: an umbrella review (a review of systematic reviews). BMC Public Health. 2021;21(1):2091. [FREE Full text] [doi: 10.1186/s12889-021-12169-7] [Medline: 34775947]

40.   Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, 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(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]

41.   Ding EY, Svennberg E, Wurster C, Duncker D, Manninger M, Lubitz SA, et al. Survey of current perspectives on consumer-available digital health devices for detecting atrial fibrillation. Cardiovasc Digit Health J. 2020;1(1):21-29. [FREE Full text] [doi: 10.1016/j.cvdhj.2020.06.002] [Medline: 32924024]

42.   Brandes A, Stavrakis S, Freedman B, Antoniou S, Boriani G, Camm AJ, et al. Consumer-led screening for atrial fibrillation: frontier review of the AF-SCREEN international collaboration. Circulation. 2022;146(19):1461-1474. [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.058911] [Medline: 36343103]

Abbreviations

AUC: area under the curve
JBI: Joanna Briggs Institute
LFD: lateral flow device
MeSH: Medical Subject Headings
PPV: positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses