Real-world accuracy of wearable activity trackers for detecting medical conditions: a systematic review and meta-analysis.

Ben Singh, Sebastien Chastin, Aaron Miatke, Rachel Curtis, Dorothea Dumuid, Jacinta Brinsley, Ty Ferguson, Kimberley Szeto, Catherine Simpson, Emily Eglitis, Iris Willems, Carol Maher

Table of Contents

Real-world accuracy of wearable activity trackers for detecting medical conditions: a systematic review and meta-analysis.

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

[1]University of South Australia Adelaide AU
[2]Ghent University Ghent BE

Corresponding Author:
Ben Singh PhD
University of South Australia
Corner of North Terrace and, Frome Rd
Adelaide
AU

Abstract

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

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

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

Results: 28 studies were included, involving a total of 1,226,801 participants (mean age range: 28.6 to 78.3). 16 studies (57.1%) used wearables for diagnosis of COVID-19, 5 (17.9%) for atrial fibrillation, 3 (10.7%) for arrhythmia or abnormal pulse, 3 (10.7%) for falls and 1 (3.6%) for viral symptoms. The devices used were Fitbit (n=6), Apple watch (n=6), Oura ring (n=3), a combination of devices (n=7), Emphatica E4 (n=1), Dynaport MoveMonitor (n=2), Samsung Galaxy Watch (n=1) and other/not specified (n=2). COVID-19 detection: Meta-analyses showed a pooled AUC 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%). Atrial fibrillation detection: Pooled PPV 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%). Falls 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 improvement are required to enhance their diagnostic precision and applicability. Clinical Trial: PROSPERO ID: CRD42023407867.

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

    Please make my preprint PDF available to anyone at any time (recommended).

    Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

    Only make the preprint title and abstract visible.

✔ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

TITLE: Real-world accuracy of wearable activity trackers for detecting medical conditions: a systematic review and meta-analysis.

Authors: Ben Singh[1], Sebastien Chastin[2,3], Aaron Miatke[1], Rachel Curtis[1], Dorothea Dumuid[1], Jacinta Brinsley[1], Ty Ferguson[1], Kimberley Szeto[1], Catherine Simpson[1], Emily Eglitis[1], Iris Willems[4], Carol Maher[1].

[1] Alliance for Research in Exercise Nutrition and Activity (ARENA), University of South Australia.

[2] School of Health and Life Sciences, Glasgow Caledonian University, UK; Department of Movement and Sports Sciences, Ghent University, 9000, Ghent, Belgium.

[4] Department of Movement and Sports Sciences, Ghent University, 9000 Ghent, Belgium.

[3] Department of Rehabilitation Sciences, Ghent University, 9000 Ghent, Belgium.

Corresponding author

Dr. Ben Singh

*   Alliance for Research in Exercise, Nutrition and Activity (ARENA), University of South Australia, Adelaide, SA, 5001, Australia

8   ben.singh@unisa.edu.au

Summary

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

Objective: The aim of this systematic review and meta-analysis was to evaluate whether wearable activity trackers can be used to detect disease and medical events.

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

Results: 28 studies were included, involving a total of 1,226,801 participants (mean age range: 28.6 to 78.3). 16 studies (57.1%) used wearables for diagnosis of COVID-19, 5 (17.9%) for atrial fibrillation, 3 (10.7%) for arrhythmia or abnormal pulse, 3 (10.7%) for falls and 1 (3.6%) 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/not specified (n=2). *COVID-19 detection:* Meta-analyses showed a pooled AUC 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%). *Atrial fibrillation detection:* Pooled PPV 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%). *Falls detection:* Pooled sensitivity was 81.9% (95% CI: 75.1-88.1%) and specificity was 62.5% (95% CI: 14.4-100%).

Conclusion: 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.

Systematic review registration: PROSPERO ID: CRD42023407867.

INTRODUCTION

As healthcare budgets around the world continue to soar, the need for cost-effective interventions that both reduce healthcare 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 behaviours, 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 for intervening on daily activity patterns and downstream health outcomes [7-12]. These studies have shown that interventions utilising 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 like 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, Alban-Cadena et al.'s [15] systematic review evaluated wearable sensors for monitoring Parkinson's disease-related gait impairments and symptoms like tremor, bradykinesia, and dyskinesia. However, most included studies were very small (10-20 participants) and were conducted in controlled laboratory environments using specialised setups like multi-sensor accelerometer arrays worn on the ankles and spine. While offering potential for home-based rehabilitation, the generalisability 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 generalisability 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 like arrhythmias, atrial fibrillation, myocardial infarction, and heart failure [16, 17]. Lee et al.'s [18] meta-analysis 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 aims 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, healthcare providers, policymakers, 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 manuscript is reported according to PRISMA [20] guidelines.

Selection criteria and search strategy

The inclusion criteria are summarised in Supplementary content 1. The inclusion criteria were developed using the population, exposure, outcomes and study type (PEOS) criteria as follows: Population: adult population (aged 18 years or older) in free living conditions, that have not been recruited based on a specific health condition or diagnosis; Exposure: use of a wearable activity tracker (e.g., Fitbit, Apple Watch or a research-grade accelerometer) for the detection of any disease or medical event (e.g., atrial fibrillation, the onset of infectious disease, falls). To be eligible, the wearable activity tracker had to be able to detect movement behaviour (i.e., include an accelerometer), but could also include other types of sensors (e.g., light sensor, temperature sensor). The wearable activity tracker had to consist of a single device worn on a single body location (e.g., on the wrist or chest, not across both); Outcomes: studies needed to assess actual diagnosis of a medical condition or occurrence of events that had clinical relevance (e.g., falls). Eligible studies needed to report an outcome related to diagnostic accuracy, such as specificity and/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% confidence intervals (CIs). Study type: Validation studies conducted under free-living conditions that were reported in a peer-reviewed journal article 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 two 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 healthcare providers or researchers as part of a formal monitoring program, and the detection of a specific clinical event or disease was a pre-specified outcome measure of the study. Studies examining consumer-driven self-tracking with personal wearables outside of a healthcare 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 which cannot track activity levels (e.g., continuous glucose monitors); studies evaluating an array of wearable sensors worn at multiple body locations (e.g., watch plus skin patch) or pedometers; studies measuring the association between an exposure and an outcome (for example, using odds ratios, relative risk, hazard ratios); lab- or hospital-based studies; and conference abstracts or dissertations.

Ten databases were searched (CINAHL, The 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 Medical Subject Headings (MeSH) term searches for terms related to "wearable device" and "detection" (see Supplementary content 2 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 pre-defined diagnostic or event outcomes. Database searches were limited to peer-reviewed journal articles published in English-language from inception to 1 April 2023.

Data management and extraction

Search results were imported into ASReview (Version 2.0, ASReview Community, Utrecht, Netherlands), an open-source software artificial intelligence (AI) tool designed for screening studies for systematic reviews. Title/abstract screening was conducted in ASReview (Version 2.0, ASReview Community, 2023) by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM or EE). The software employs an active learning algorithm that iteratively selects the most relevant studies for inclusion based on the initial judgments made by the research team. Screening was stopped when 100 consecutive non-relevant articles were screened. Following title/abstract screening, results were then imported to EndNote x9 (Clarivate, Philadelphia, PA) where duplicates were removed, then exported into Covidence (Veritas Health Innovation, Melbourne, Australia) 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 standardised extraction form in Covidence. The risk of bias of the included reviews was assessed by two independent reviewers in duplicate using the Joanna Briggs Institute

(JBI) Critical Appraisal Checklist for Diagnostic Test Accuracy Studies. Studies were rated out of 9 for the following items: 1) enrolment 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 pre-specification 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 i) COVID-19 detection, ii) atrial fibrillation or arrhythmia detection, and iii) falls detection. Outcomes of interest were analysed and data were pooled using sensitivity (%), specificity (%), area-under-curve (AUC, %), accuracy (%) and positive predictive value (PPV), with 95% confidence intervals (CIs) as the effects measures. Sensitivity (%) denotes the percentage of individuals with the disease/condition correctly identified by the test, while specificity (%) represents the percentage of those without the disease/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/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's 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 other) 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 (v16, Stata Corp, College Station, TX, USA).

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 (i.e., individual RCTs); B:

consistent level 2 (i.e., individual cohort studies) or 3 studies (i.e., individual case-control studies) or extrapolations from level 1 studies; C: level 4 studies (i.e., case series) or extrapolations from level 2 or 3 studies; or D: level 5 (i.e., 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 type of wearable for atrial fibrillation and fall detection, due to an insufficient number of studies.

RESULTS

Of the 21,429 records identified following the database search, 28 were eligible (see Figure 1 for PRISMA flowchart including reasons for exclusions; see Supplementary content 3 for a complete list of full texts that were excluded during the final stage of screening, with reasons). An overview of all included studies characteristics is shown in Table 1. There was a total of 1,226,801 participants (median [range] sample size: 264 [29 to 455,699]). Mean participant age ranged between 28.6 and 78.3 (median=43.7) years and 21 (75%) studies involved female and male participants (gender was not reported in 7 [25%] studies). Sixteen studies (57.1%) evaluated COVID-19, five (17.9%) studies evaluated atrial fibrillation, three (10.7%) studies assessed a broad range of cardiac arrhythmias, three (10.7%) assessed falls and one (3.6%) 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 (i.e., 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/not specified (n=2). Median score for the JBI Critical Appraisal Checklist for Diagnostic Test Accuracy Studies was 6 (range: 1 to 9) out of 9 (Supplementary content 4).

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, %), (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 nine studies showed a pooled AUC of 80.15% (95%: 71.03%-89.27%) and five studies had a pooled accuracy of 87.54% (95% CI: 81.57%-93.51%). Pooled sensitivity from eight studies was 79.53% (95% CI: 67.73%-91.33%), and seven 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 Supplementary content 5 and 6 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 82.69% and 74.62% (respectively).

Results of sensitivity analyses are shown in Supplementary content 7. Following 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.

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).

Results of sensitivity analyses are shown in Supplementary content 8. Following 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.

Falls detection

Meta-analysis results of sensitivity and specificity for falls detection are shown in Figure 5. Meta-analyses of 2 studies showed a specificity of 62.54% (95% CI: 14.43%-100%) and 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 falls detection.

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

DISCUSSION

In this study, we set out to systematically review and meta-analyse 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 focussed 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 falls 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 towards 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 PCR tests, widely employed, 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 like sleep apnoea 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 focussed on 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 (e.g., [24, 25]) suggesting that a wider range of diagnostic outcomes may become available in future. Furthermore, many studies were excluded because they focused on monitoring of symptoms in people with a known diagnosis (for example, seizures in people with epilepsy [26], and freezing gait in Parkinson's disease [27]) which was outside the scope of our study, but highlights wearable activity trackers' potential for medical condition monitoring.

Our 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 100.0% [28]. Comparatively, our study's pooled sensitivity for wearable-detected COVID-19 was 79.5% (range: 51.3-100%), which is in line with the LDF 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 favourably to other clinical tests. For example, the sensitivity and specificity of 12-lead electrocardiogram for detecting atrial fibrillation has previously 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 (FDA) 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 sensor. 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's 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.

Our 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 analysed a robust dataset from 28 studies, involving over one 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 usage patterns. While this combination enhances the generalisability 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 (e.g., 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 electrocardiogram. 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 healthcare by giving them immediate feedback and data about their condition. They may also contribute to improved surveillance and resource planning for healthcare systems, which could be particularly useful in time 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, FDA-approved devices relied on optical heart rate sensors providing photoplethysmography data, capable of identifying irregular heart rhythms characteristic of arrhythmias. Fall detection primarily utilizes accelerometer data, with wrist-worn placement crucial for sensing sudden deceleration and impact forces. However, false positives persist due to non-fall 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 like 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, healthcare professionals have reported an uptick in patient consultations triggered by atrial fibrillation alerts from wearables, resulting in a surge of medical tests such as electrocardiogram to confirm diagnoses [41]. While some clinicians see this as an advancement in patient-initiated healthcare, 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 future, our meta-analysis could be updated to provide insight into the accuracy of such diagnostics by condition, device and population.

Conclusions

Our 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 like 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 towards COVID-19, leaving a significant gap in our understanding of how these devices can be utilised 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 realise their full potential as tools for healthcare monitoring and intervention.

Table 1 Overview of characteristics of included studies (n=28).

Author, year	Country	Sample size, gender	Age	Wearable device	Data Input/ Vital Signs Collected	Disease or medical event	Data Labelling/Data Segmentation	Reference standard	Algorithm or statistical model	Algorithm/ accelerometer data open or available to access?
Abir 2022	Qatar	120 Gender: NR	NR	Fitbit	Heart rate, steps	COVID-19	At least 20 days before the symptom onset to 21 days afterward.	Reverse transcription-polymerase chain reaction	Long Short-term Memory Variational Autoencoder-based anomaly detection framework	No /Yes
Alavi 2022	United States	3,318 F: 1,854 (56%) M: 1,464 (44%)	Median (range): 44 (19-79)	Fitbit, Apple Watch & Garmin	Heart rate, steps	COVID-19	3 days before symptom onset, for the next 15 days.	COVID-19 diagnosis for asymptomatic cases and self reported symptom for symptomatic individuals	NightSignal, RHRAD, CuSum	Yes /Yes
Caillol 2021	France	256 F: 108 (42%) M: 148 (58%)	66±6	Apple Watch	Sinus rhythm	Arrhythmia	NR	12-lead ECG	Cohen's kappa	No /No
Cleary 2022	United States	105 F: 53 (51%) M: 52 (49%)	28.6±2.8	Fitbit & Apple watch	Heart rate, steps and sleep	COVID-19	0–7 days after symptom onset as test periods.	COVID-19-positive test	RHRmetric, SLEEPmetric, STEPmetric, SENSORmetric	Yes /Yes
D'Haese 2021	United States	867 F: 313 (36.1%) M: 236 (27.2%) Unknown: 318 (36.7%)	37.6±11.3	Oura Ring	Body temperature, sleep, activity, heart rate, respiratory rate, heart rate variability	Viral symptoms	3 days before symptom onset	Self-reported viral illness	Labelling model and forecasting model	Yes /No
Gadaleta 2021	United States	38,911 F: 23,736 (61%) M: 15,175 (39%)	NR	Fitbit & Apple Watch	Heart rate, sleep, activity	COVID-19	NR	COVID-19 nasal swab test	CatBoost gradient boosting	Yes /Yes
Guo 2019	China	187,912 F: 24,938 (13.3%) M: 16,2974 (86.7%)	34.7±11.5	Honor Band 4, Huawei Watch & Honor Watch	Pulse rhythm	Atrial Fibrillation	NR	Clinical evaluation, ECG, or 24-h Holter monitoring	Kruskal-Wallis test and a photoplethysmography algorithm	Yes /No
Hassantabar 2012	Italy	87 Gender: NR	NR	Empatica E4	Galvanic skin response, temperature, interbeat interval, oxygen saturation	COVID-19	NR	COVID-19-positive test	Naïve Bayes, Random Forest, Ada Boost, Decision Tree, SVM, k-NN, deep neural	Yes /No

Author	Country	Participants	Age	Device	Measures	Condition	Symptom definition	Reference standard	Analysis	Col1/Col2
									network model with grow-and-prune synthesis	
Hirten 2021	United States	297 F: 204 (69.4%) M: 93 (31.3%)	36.3±9.8	Apple watch	Heart rate variability	COVID-19	Defined being symptomatic as the first day of a reported symptom	COVID-19-positive nasal swab PCR test	Mixed-effect Cosinor model.	No /No
Liu 2021	Europe	87 F: 24 (27.6%) M: 63 (72.4%)	46.5±10.5	Fitbit	Heart rate	COVID-19	NR	COVID-19 diagnosis	Conventional neural network, Multilayer Perceptrons, Long Short-Term Memory Networks, conventional convolutional auto-encoder, contrastive conventional convolutional auto-encoder	No /No
Lockhart 2021	United States	171 Gender: NR	Fallers: 75.4±8.7 Non-fallers: 75.7±7.6	Inertial measurement unit (accelerometer worn on sternum)	Trunk kinematics	Falls	NR	Ten-meter walk test and fall history	Modified continuous wavelet transform method	Yes /No
Lonini 2020	United States	29 F: 11 (50%) M: 11 (50%)	42.9±15.9	Soft wearable sensor (accelerometer worn on suprasternal notch)	RR intervals, steps, RR and frequency spectrum of cough signals	COVID-19	Labelled snapshots as COVID-19 positive and negative	COVID-19 diagnosis PCR test	Logistic Regression (Supervised machine learning)	No /No
Lubitz 2022	United States	455,699 F: 323,365 (71%) M: 132,334 (29%)	Median (IQR): 47 (35–58)	Fitbit	Heart rate	Atrial Fibrillation	NR	Single-lead ECG patch monitor	Fitbit PPG RhythmDetect Software System algorithm	No /No
Mason 2022	United States	63,153 F: 24,374 (40%) M: 36,632 (60%) Other: 56 (<0.1%)	18-30 y: n=8,555 (14%) 31-40 y: n=16,756 (27%) 41-50y: n=17,502 (29%) 51-80y: n=18,148	Oura Ring	Heart rate, heart rate variability, dermal temperature, respiratory rate, physical activity (METs)	COVID-19	NR	COVID-19 PCR test	Machine learning, random forest models	No /No

			(30%) 81+ y: n=102 (0.2%)							
Mishra 2020	United States	5,262 32 diagnosed F: 25 (78.1%) M: 7 (21.9%)	Mean (range): 47 (27-67)	Fitbit	Heart rate, steps	COVID-19	Dates of symptom onset and diagnosis to define sick periods	COVID-19 diagnosis	RHR-Diff, HROS-AD, CuSum	Yes /Yes
Natarajan 2020	United States	2,745 Gender: NR	40.7±12.3	Fitbit	RR, heart rate, and HRV	COVID-19	Data from 2nd to 6th day of symptom onset labelled as sick	COVID-19 diagnosis PCR test	CNN (Supervised deep learning)	No /No
Nestor 2023	Canada	6,926 F: 6,012 (86.8%) M: 914 (13.2%)	NR	Fitbit	Night-time RR, heart rate, HRV and symptom report	COVID-19	Days between self reported symptom onset and self-reported recovery labelled as positive	COVID-19 diagnosis	XGBoost, XGBoost and GRU-D	No /Yes
Perez 2019	United States	419,297 F: 177,087 (42%) M: 238,700 (57%) Other: 396 (0.1%) NR: 3,114 (0.7%)	41±13	Apple Watch	Pulse rate	Atrial fibrillation	NR	ECG patch	Irregular Pulse Notification Algorithm	No /No
Ploux 2021	France	260 F: 109 (42%) M: 151 (58%)	66±6	Apple Watch	Sinus rhythm	Arrhythmia	NR	12-lead ECG	NR	No /No
Quer 2021	United States	30,529 F: 18,922 (62%) M: 11, 607 (38%)	NR	Fitbit & Apple Watch	Heart rate, sleep, activity	COVID-19	First date of symptoms to seven days after symptoms considered infectious	COVID-19 diagnosis	RHRMetric, SleepMetric, ActivityMetric, SymptomMetric, SensorMetric, OverallMetric	Yes /Yes
Skibińska 2021	Czech Republic	54 Gender: NR	NR	Fitbit, Apple Watch & various others	Heart rate, steps, sleep	COVID-19	NR	COVID-19 diagnosis	XGBoost, k-NN, SVM, Logistic Regression Decision Tree Random Forest	Yes /Yes
Skibińska 2022	Czech Republic	58 Gender: NR	NR	Fitbit, Apple Watch & various others	Heart rate, steps, sleep	COVID-19	NR	COVID-19 diagnosis	XGBoost, k-NN, Logistic Regression, Support Vector Machine (SVM), Decision Tree, and	Yes /Yes

Study	Country	Sample	Age	Device	Measures	Condition	Onset	Reference standard	Analysis	Yes/Yes
									Random Forest.	
Smarr 2020	United States	50 F: 22 (44%) M: 28 (66%)	43.7±11.0	Oura ring	Temperature, heart rate, heart rate variability, and respiration rate	Fever and COVID-19	Onset of reporting of symptoms	COVID-19 diagnosis	Wavelet analysis	Yes /Yes
Tison 2018	United States	9,750 M: 5395 (55.3%) F: 2195 (22.5%) Unknown: 973 (10%) NR: 1187 (12.2%)	Atrial Fibrillation: 55.7±14.2 No Atrial Fibrillation: 41.5 ±11.9	Apple watch	Heart rate, step count, R-R interval	Atrial Fibrillation	NR	12-lead ECG	Deep neural network development and training	Yes /No
van Schooten 2016	Netherlands	319 F: 161 (50.5%) M: 158 (49.5%)	75.5±6.9	Dynaport MoveMonitor accelerometer	Daily-life gait quality	Falls	NR	Self-reported falls (telephone survey and diary)	Principal component analysis	No /No
Wyatt 2020	United States	264 F: 162 (61.4%) M: 102 (38.6%)	Median (IQR): 55 (37.7-69.0)	Apple Watch	Heart rate	Abnormal pulse	NR	Electronic health record data of 12-lead ECG, Holter monitor and chest x-ray.	Chi-square tests.	No /No
Weiss 2013	United States	71 F: 46 (64.8%) M: 25 (35.2%)	78.3±4.7	DynaPort accelerometer	Gait Quality	Falls	NR	Self-reported falls	Binary logistic regression	No /No
Zhu 2022	United States	204 Gender: NR	NR	Samsung Galaxy Watch	Photoplethysmography	Atrial Fibrillation	NR	One-channel ECG (chest patch)	Sample-wise signal quality estimator, hybrid decision model, high-level decision generator, sensor contact monitor	No /No

ECG: Electrocardiogram.
F: Female.
IQR: Interquartile range.
M: Male.
NR: Not reported.

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: 17-25.

2.   Wang W, Yan Y, Guo Z, 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 Journal 2021; 12(4): 403-433.

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): 913-925.

4.   Shin G, Jarrahi MH, Fei Y, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. Journal of Biomedical Informatics 2019; 93: 103153.

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 Science, Medicine and Rehabilitation 2015; 7(1): 24.

6.   Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity 2015; 12(1): 159.

7.   Ferguson T, Olds T, Curtis R, 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.

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.

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 Network Open 2023; 6(6): e2318478-e2318478.

10.   Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. British Journal of Sports Medicine 2021; 55(8): 422-432.

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

20

cancer survivors: A systematic review and meta-analysis. Journal of Sport and Health Science 2022; 11(2): 184-193.

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 & Research 2019; 71(6): 758-767.

13.   Bayoumy K, Gaber M, Elshafeey A, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nature Reviews Cardiology 2021; 18(8): 581-599.

14.   Shei R-J, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers–advanced technology or advanced marketing? European Journal of Applied Physiology 2022; 122(9): 1975-1990.

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. Journal of Medical Engineering & Technology 2021; 45(7): 532-545.

16.   Pay L, Yumurtaş A, Satti DI, et al. Arrhythmias Beyond Atrial Fibrillation Detection Using Smartwatches: A Systematic Review. Anatol J Cardiol 2023; 27(3): 126-131.

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.

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.

19.   Cheong SHR, Ng YJX, Lau Y, Lau ST. Wearable technology for early detection of COVID-19: A systematic scoping review. Preventive Medicine 2022; 162: 107170.

20.   Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372: n71.

21.   Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. Computers in Biology and Medicine 2000; 30(3): 127-134.

22.   Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

21

23. Oxford Centre for Evidence-Based Medicine. Levels of evidence (March 2009). University of Oxford, 2022. Available from https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009.

24. Johansson D, Ohlsson F, Krýsl D, et al. Tonic-clonic seizure detection using accelerometry-based wearable sensors: A prospective, video-EEG controlled study. Seizure 2019; 65: 48-54.

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): 045.

26. Kusmakar S, Karmakar C, 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.

27. Ahlrichs C, Samà A, Lawo M, 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.

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.

29. UK Government. (19 October 2021). Technologies Validation Group: Using tests to detect COVID-19. Retrieved from https://www.gov.uk/guidance/technologies-validation-group-using-tests-to-detect-covid-19

30. Welton NJ, McAleenan A, Thom HH, et al. Screening strategies for atrial fibrillation: a systematic review and cost-effectiveness analysis. Health Technol Assess 2017; 21(29): 1-236.

31. Lærke Marius K, Nicklas V, Grazina U, Jes Sanddal L, Axel Cosmus Pyndt D, Lars F. Diagnostic accuracies of screening for atrial fibrillation by cardiac nurses versus radiographers. Open Heart 2019; 6(1): e000942.

32. Google. New Fitbit feature makes AFib detection more accessible. Google Blog. Retrieved from https://blog.google/products/fitbit/irregular-heart-rhythm-notifications/. Published April 11, 2022. Accessed October 3 2023.

33. Kritz F. Apple Watches Have Two New FDA-Cleared Health Applications. Retrieved from https://array.aami.org/content/news/apple-watches-have-two-new-fda-cleared-health-applications. Published June 29, 2022. Accessed October 3 2023.

34. Malik A. Garmin launches a new FDA-cleared ECG app for the Venu 2 Plus. Retrieved from https://techcrunch.com/2023/01/24/garmin-new-fda-cleared-ecg-app-venu-2-

22

plus/#:~:text=With%20the%20new%20ECG%20App,more%20insight%20into%20your %20health. Published January 25, 2023. Accessed October 3 2023.

35.   Samsung Newsroom US. Retrieved from https://news.samsung.com/us/fda-cleared-irregular-heart-rhythm-notification-for-galaxy-watch/2023. Published May 8, 2023. Accessed October 3 2023.

36.   Natarajan A, Su H-W, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. npj Digital Medicine 2020; 3(1): 156.

37.   Hassantabar S, Stefano N, Ghanakota V, et al. CovidDeep: SARS-CoV-2/COVID-19 Test Based on Wearable Medical Sensors and Efficient Neural Networks. IEEE Transactions on Consumer Electronics 2021; 67(4): 244-256.

38.   Chen M, Wang H, Yu L, et al. A Systematic Review of Wearable Sensor-Based Technologies for Fall Risk Assessment in Older Adults. Sensors. 2022; 22(18): 6752.

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.

40.   Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. European Heart Journal 2021; 42(5): 373-498.

41.   Ding EY, Svennberg E, Wurster C, et al. Survey of current perspectives on consumer-available digital health devices for detecting atrial fibrillation. Cardiovascular Digital Health Journal 2020; 1(1): 21-29.

42.   Brandes A, Stavrakis S, Freedman B, et al. Consumer-led screening for atrial fibrillation: frontier review of the AF-SCREEN International Collaboration. Circulation 2022; 146(19): 1461-1474.
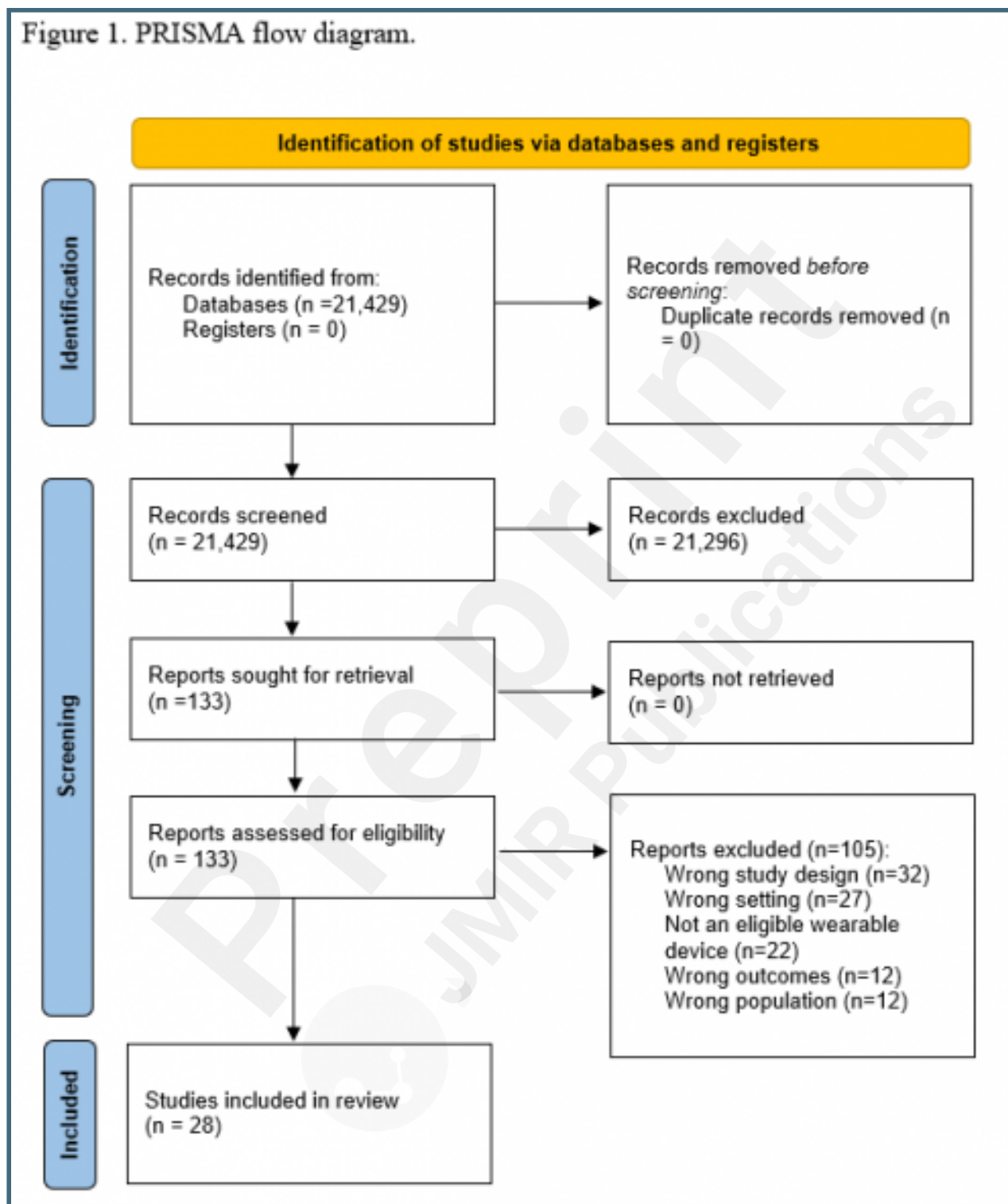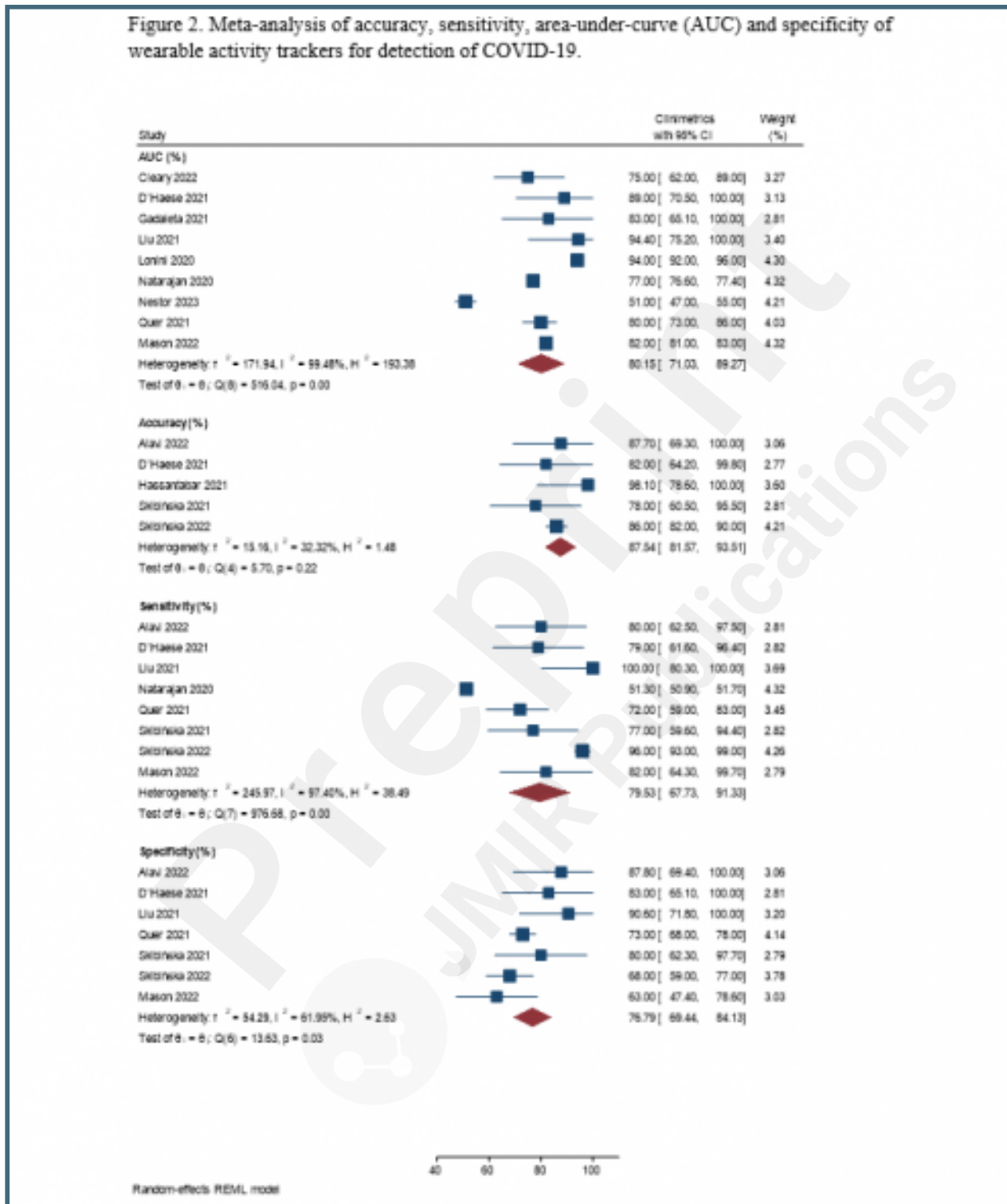
Supplementary Files
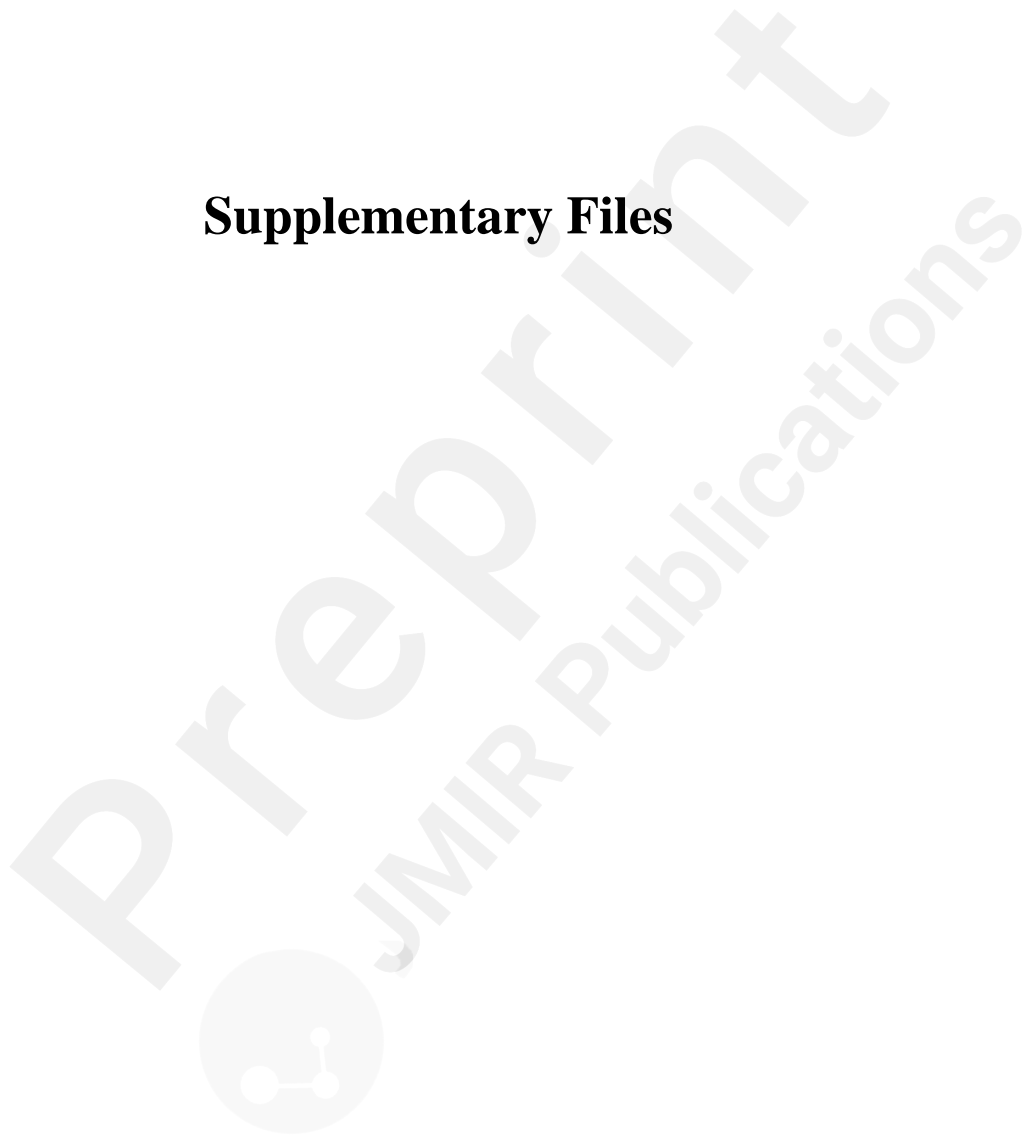
Figures

PRISMA flow diagram.
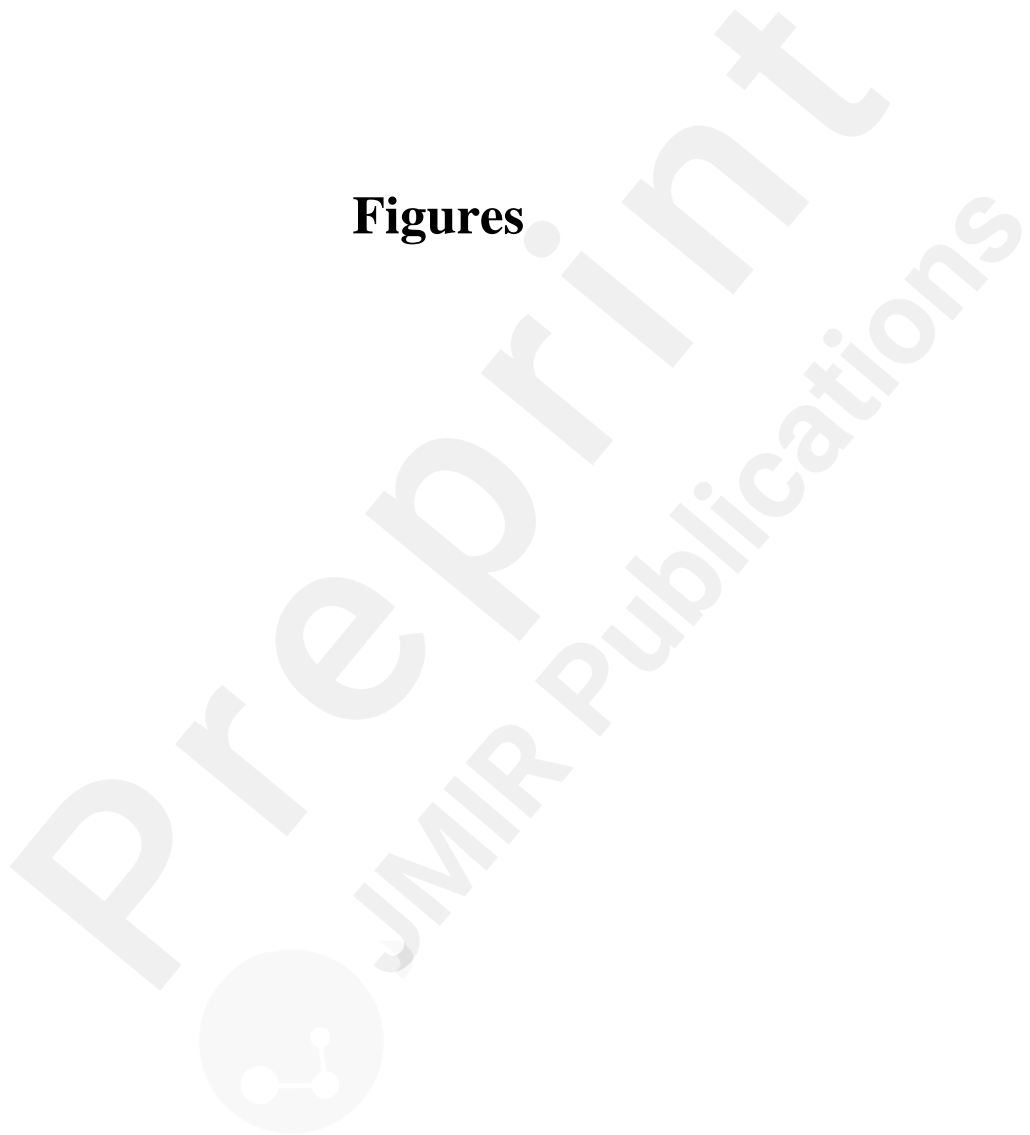
Figure 1. PRISMA flow diagram.

Meta-analysis of accuracy, sensitivity, area-under-curve (AUC) and specificity of wearable activity trackers for detection of COVID-19.
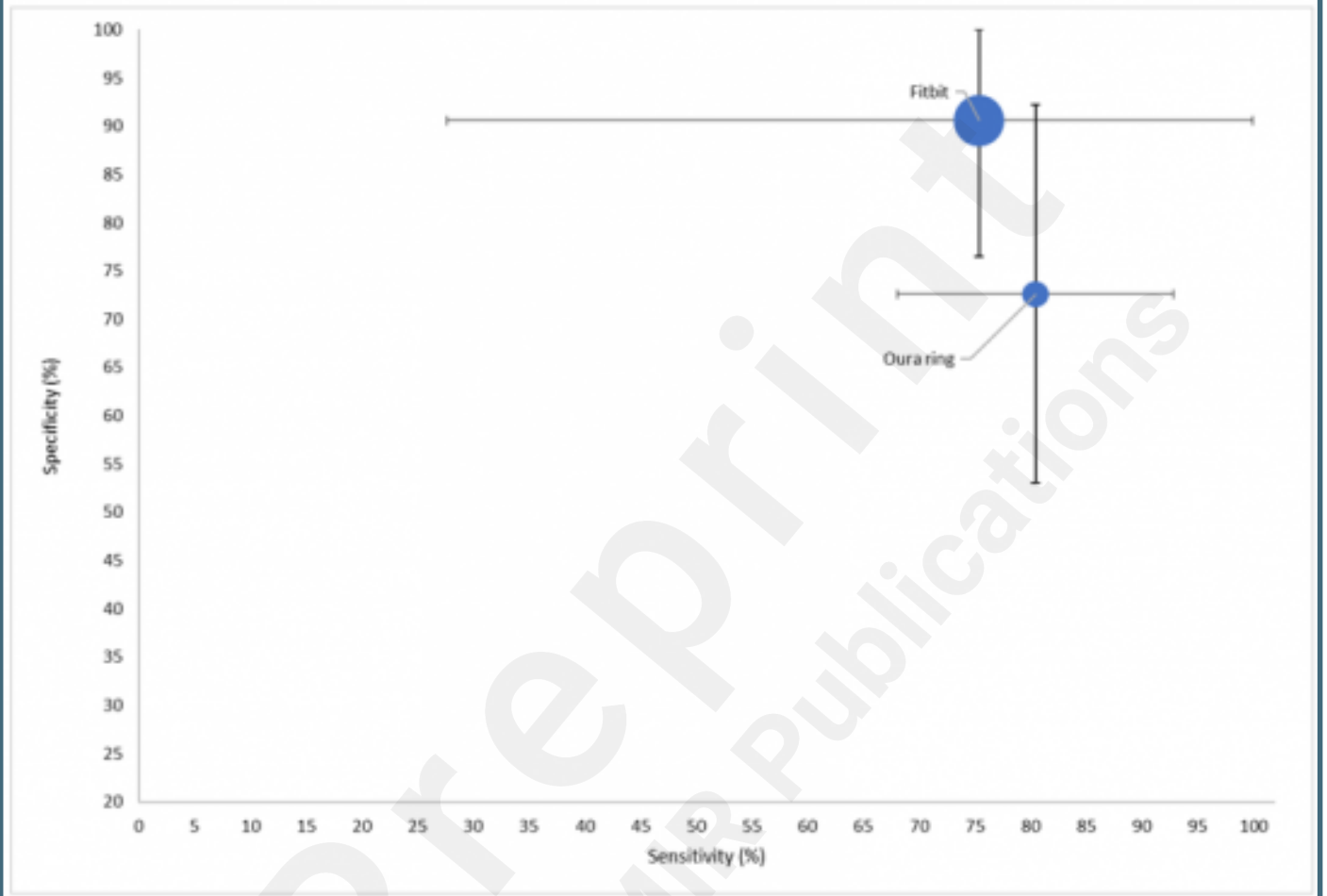

Figure 2. Meta-analysis of accuracy, sensitivity, area-under-curve (AUC) and specificity of wearable activity trackers for detection of COVID-19.
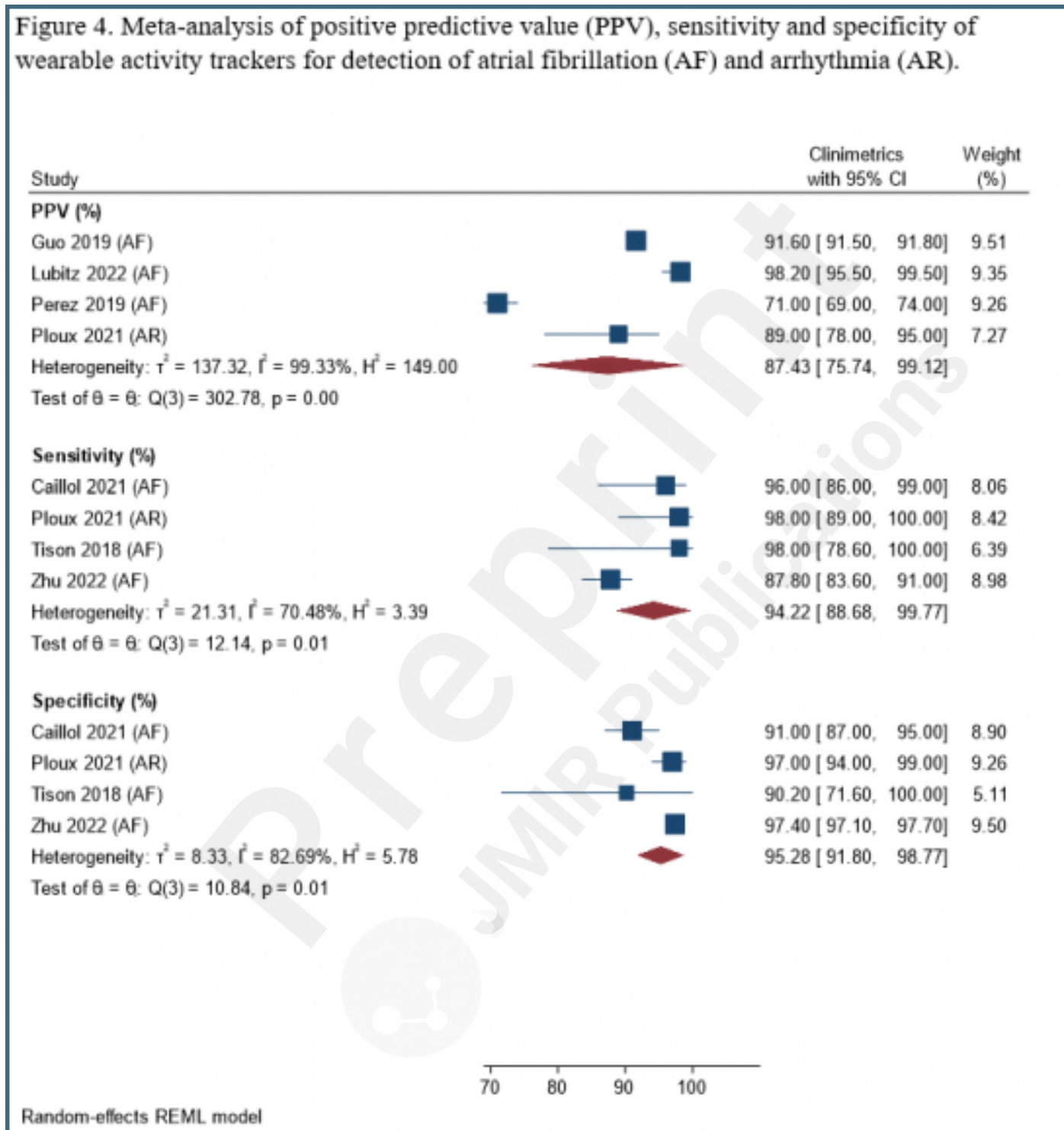
Overview of sensitivity and specificity for the different devices for COVID-19 detection.

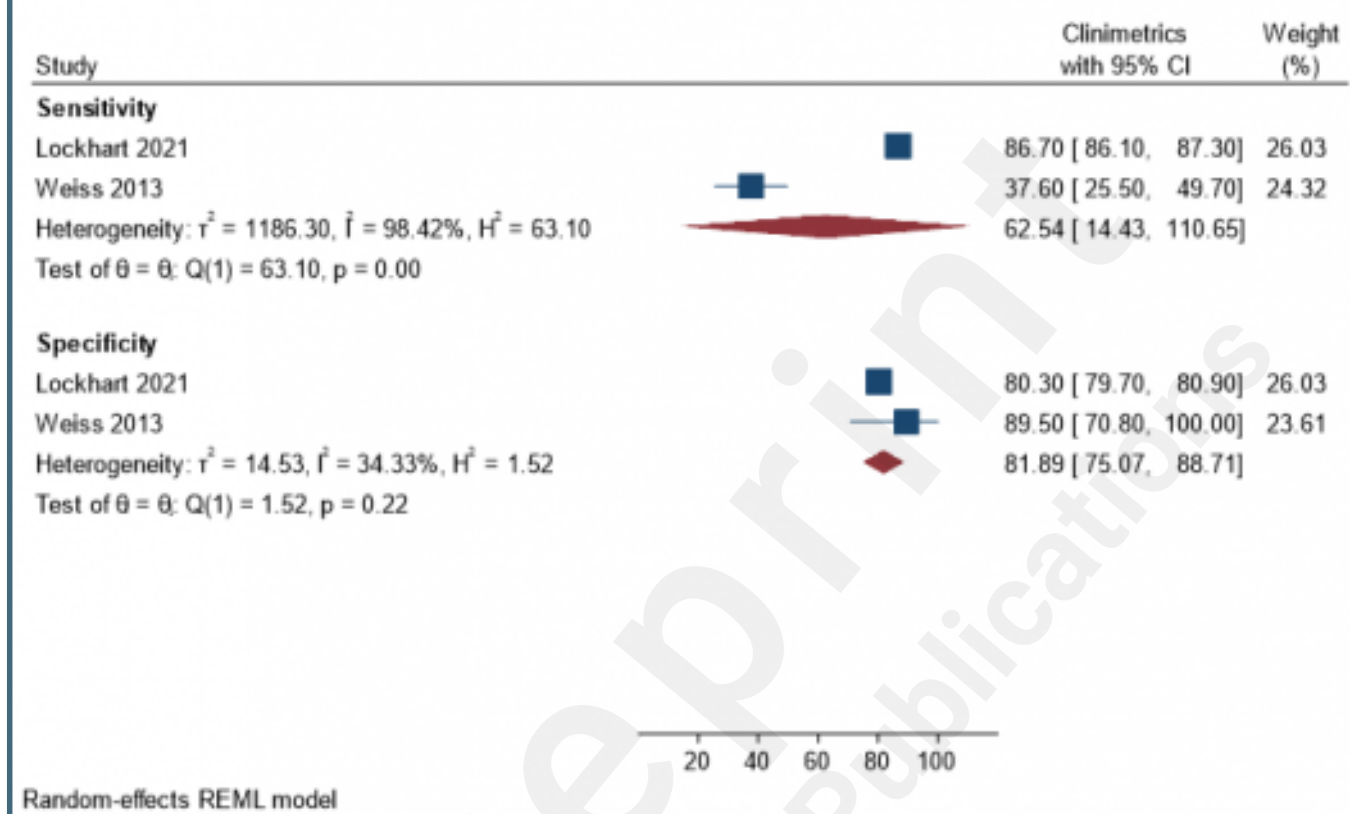Figure 3. Overview of sensitivity and specificity for the different devices for COVID-19 detection.

Meta-analysis of positive predictive value (PPV), sensitivity and specificity of wearable activity trackers for detection of atrial fibrillation (AF) and arrhythmia (AR).

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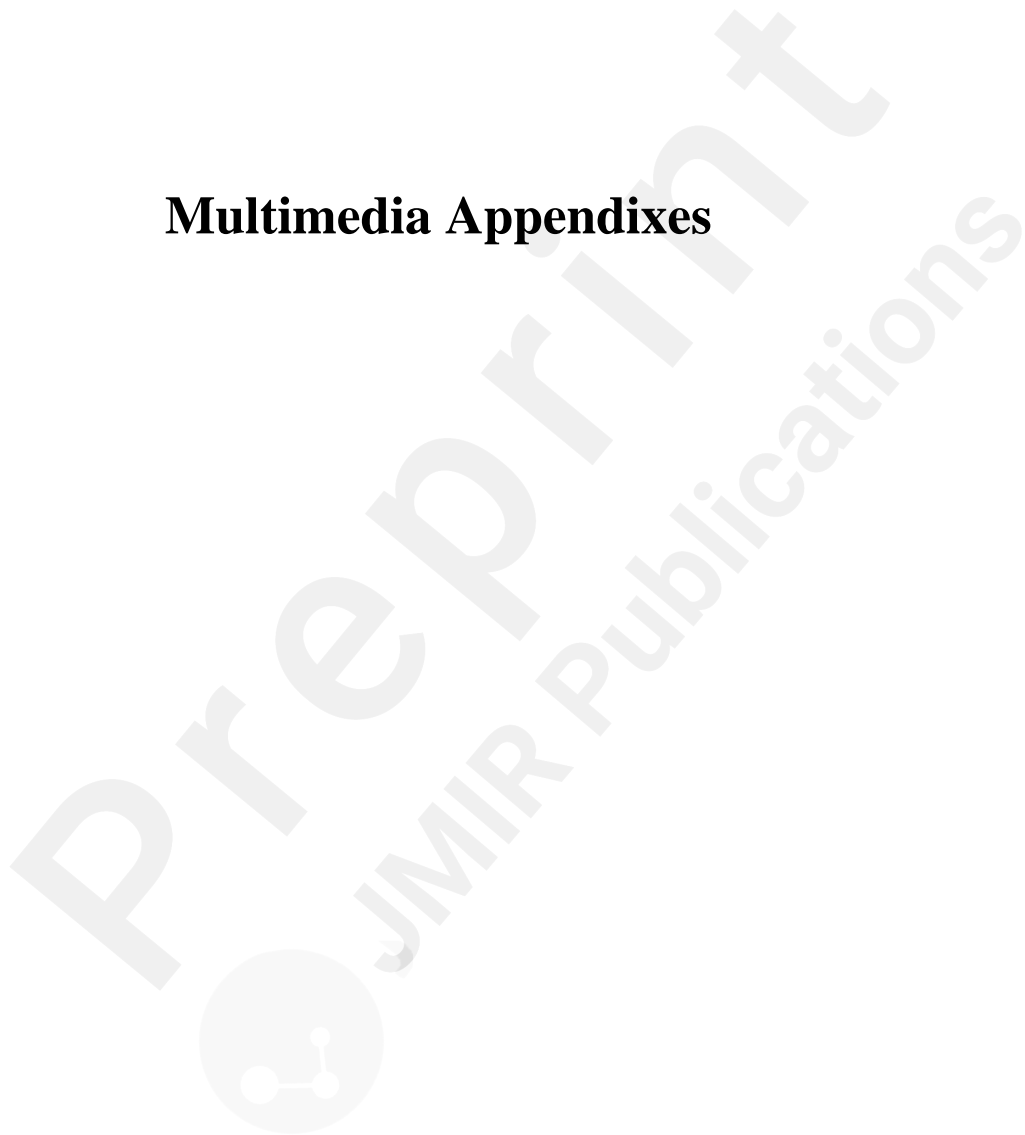
Meta-analyses of sensitivity and specificity of wearable activity trackers for detection for falls detection.

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

Multimedia Appendixes

Supplementary material.
URL: http://asset.jmir.pub/assets/7698ea2c548da7d0e536a64c49862ace.docx

CONSORT (or other) checklists

PRISMA Checklist.
URL: http://asset.jmir.pub/assets/5ecc3f0af12e9c4f8ba5e615cd9f1469.pdf