Baltic Journal of Health and Physical Activity

Volume 14 | Issue 4 Article 8

2022

Health outcomes of Fitbit, Garmin or Apple Watch-based interventions: A systematic review of systematic reviews

Hossein MOTAHARI-NEZHAD
Doctoral School of Business and Management, Corvinus University of Budapest, Budapest, Hungary,
h.motahari.lib@gmail.com

Fateme ZARE
University of Physical Education, Budapest, Hungary

Hadi AKBARI
Department of Sport Sciences, Faculty of Literature and Humanities, University of Zabol, Zabol, Iran

Aslan SADEGHDAGHIGHI
Semmelweis University, Budapest, Hungary

Recommended Citation

Motahari-Nezhad H, Zare F, Akbari H, Sadeghdaghighi A. Health outcomes of Fitbit, Garmin or Apple Watch-based interventions: A systematic review of systematic review. Balt J Health Phys Act. 2022;14(4):Article8. https://doi.org/10.29359/BJHPA.14.4.08

Health outcomes of Fitbit, Garmin or Apple Watch-based interventions: A systematic review of systematic reviews

Abstract

Introduction: A rising number of systematic reviews of smartwatch clinical effectiveness have been published; however, the findings are conflicting and need further investigation. Our purpose was to assess the methodological and evidence synthesis quality of meta-analyses of the therapeutic efficacy of Fitbit, Garmin, or Apple Watch-based interventions. Material and Methods: Systematic searches were carried out with three electronic databases from inception to October 2021. Full text systematic reviews originally published in English in peer-reviewed journals describing meta-analyses of all clinical outcomes of Fitbit-, Garmin-, or Apple Watch -based interventions were eligible for this study. The Amstar-2 scale and GRADE were used to assess the methodological and evidence synthesis quality, respectively. Results: One study with low methodological quality (overall score of 13.5) was identified, which contained five outcomes. None of the outcomes met the GRADE criteria for high-quality evidence (0%). Two outcomes (40%) had moderate-quality evidence synthesis, while three outcomes (60%) had low-quality evidence synthesis. The most prevalent downgrading GRADE factor was a risk of bias (100%), followed by imprecision (80%), and publication bias (60%). Conclusion: Despite the widespread adoption of smartwatches, there is a large gap in the literature as there is limited evidence. This review suggests that further clinical trials and meta-analyses are needed and it formulates research recommendations.

Keywords

Fitbit, Garmin, Apple watch, Smartwatch, Amstar-2, GRADE

Baltic Journal
of Health and Physical Activity
e-ISSN 2080-9999

Review

Health outcomes of Fitbit, Garmin or Apple Watch-based interventions: A systematic review of systematic reviews

Hossein MOTAHARI-NEZHAD[1]*, Fateme ZARE[2], Hadi AKBARI[3], Aslan SADEGHDAGHIGHI[4]

[1] Doctoral School of Business and Management, Corvinus University of Budapest, Budapest, Hungary, ORCID 0000-0002-1028-4460
[2] University of Physical Education, Budapest, Hungary, ORCID 0000-0001-5726-0194
[3] Department of Sport Sciences, Faculty of Literature and Humanities, University of Zabol, Zabol, Iran, ORCID 0000-0001-7899-5528
[4] Semmelweis University, Budapest, Hungary, ORCID 0000-0001-9982-8702

* Correspondence: Hossein Motahari-Nezhad; Fovam Ter 8 1093 Budapest, Hungary; phone: +36702097967; e-mail: h.motahari.lib@gmail.com

Abstract: Introduction: A rising number of systematic reviews of smartwatch clinical effectiveness have been published; however, the findings are conflicting and need further investigation. Our purpose was to assess the methodological and evidence synthesis quality of meta-analyses of the therapeutic efficacy of Fitbit, Garmin, or Apple Watch-based interventions. Material and Methods: Systematic searches were carried out with three electronic databases from inception to October 2021. Full text systematic reviews originally published in English in peer-reviewed journals describing meta-analyses of all clinical outcomes of Fitbit-, Garmin-, or Apple Watch -based interventions were eligible for this study. The Amstar-2 scale and GRADE were used to assess the methodological and evidence synthesis quality, respectively. Results: One study with low methodological quality (overall score of 13.5) was identified, which contained five outcomes. None of the outcomes met the GRADE criteria for high-quality evidence (0%). Two outcomes (40%) had moderate-quality evidence synthesis, while three outcomes (60%) had low-quality evidence synthesis. The most prevalent downgrading GRADE factor was a risk of bias (100%), followed by imprecision (80%), and publication bias (60%). Conclusion: Despite the widespread adoption of smartwatches, there is a large gap in the literature as there is limited evidence. This review suggests that further clinical trials and meta-analyses are needed and it formulates research recommendations.

Keywords: Fitbit, Garmin, Apple watch, Smartwatch, Amstar-2, GRADE.

Citation: Motahari-Nezhad H, Zare F, Akbari H, Sadeghdaghighi A. Health outcomes of Fitbit, Garmin or Apple Watch-based interventions: A systematic review of systematic review. Balt J Health Phys Act. 2022;14(4):Article8. https://doi.org/10.29359/BJHPA.14.4.08

Academic Editor: Aleksandra Bojarczuk

Received: October 2021
Accepted: January 2022
Published: December 2022

Publisher's Note: BJHPA stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Consumer use of wearable devices has become an increasingly strong trend over recent years [1]. According to statistics, approximately 225 million wearable devices were purchased by users worldwide in 2019 [2], rising at a compound annual growth (CAGR) of approximately 13% from 2020 to 2027 [3]. Moreover, published reports confirm that most adults in Canada [4] and Australia [5] use wearable devices. Besides, as reported by Statista, the penetration rate of wearable devices among American adults is expected to reach over 25% by 2022 [6], which shows the high popularity of these devices worldwide.

Among the most popular wearable devices are the Apple Watch, Fitbit and Garmin. Apple Watch alone accounted for 49% of the global smartwatch market in 2016. Moreover, Apple Watch had a share of more than 33% of the smartwatch market in 2021 [7]. Among the various brands of wearable devices is Fitbit, which was founded in 2007 and presently is officially part of Google company. The sale of more than 120 million devices since 2009

in over 100 countries indicates the acceptance and widespread use of Fitbit worldwide [8]. Founded in Kansas in 1989, Garmin has been making activity trackers since 2006, and became the fifth largest exporter of activity trackers in the world in 2018, with a market share of more than 5% [9].

The widespread public interest in wearable devices, especially smartwatches, has provided a new opportunity for both consumers and healthcare systems for remote and continuous monitoring, and tracking health endpoints are used as an intervention and measurement devices in clinical settings.

Tracking individuals' physical activity for assessing step counts, heart rate [10], and other parameters as an intervention [11–13] as well as in conjunction with other treatments [14, 15] has resulted in changing the population's physical behavior and in revolutionizing medical decision making [16]. In addition, smartwatches, specifically Fitbit, account for 89% of published papers, 83% of clinical trials, and 95% of NIH-funded research in biomedical studies [17]. Also, conducting a search in the clinical trial electronic database on the 25th of August 2021 yielded 704 clinical Fitbit, Garmin, or Apple watch-based studies.

Despite the rising popularity of commercial smartwatches, especially in recent years, there is still uncertainty about their validity and accuracy in evaluating measures related to physical activity [1]. In other words, the existing evidence for the efficiency of tracker-based interventions is indecisive [18]. For example, in a systematic review study, synthesis of the evidence showed that Fitbit did not provide an accurate measure of the amount of energy expended in each test condition [19]. Therefore, incorrect estimation of clinical data can have a negative impact on medical decision making, leading to adverse outcomes.

A growing number of systematic reviews have been published on the clinical effectiveness of smart watches, but the results have been conflicting [20, 21] and need to be studied further. The inconsistent results from current systematic reviews call for a more systematic assessment of the strength and quality of evidence regarding the health outcomes of smartwatch-based interventions (Fitbit, Garmin, Apple Watch). Therefore, an overview of systematic reviews has been conducted in this study to assess the current quality of evidence in the field. It is evident that a lack of assessing the quality of evidence in systematic reviews can lead to prejudiced treatment guidelines, negatively affecting medical decision-making and imposing major financial burdens on healthcare systems [22].

Various instruments have been introduced to assess the quality of evidence [23], of which the GRADE – Grading of Recommendations, Assessment, Development and Evaluations [24] has universal validity and acceptance and has been adopted by the World Health Organization (WHO) and is recommended in the Cochrane handbook [25]. GRADE categorizes the quality of evidence into four levels including high, moderate, low, and very low. High quality signifies that additional research is improbable to change certainty in the effect estimate, while, very low one indicates that there is very little certainty in the effect estimate [24]. Therefore, in this study, the intention was to assess the overall strength of evidence of systematic reviews that explain a quantitative synthesis of the impact of Fitbit, Garmin, or Apple Watch on health outcomes when compared to other interventions without a Fitbit, Garmin or Apple Watch. The study provides a far-reaching summary of the quality of the existing evidence on outcomes of interventions using Fitbit, Garmin or Apple Watch.

Accordingly, the hypotheses of this review are as follows:

1. Systematic reviews evaluating the clinical effectiveness of Fitbit-, Garmin-, or Apple Watch-based interventions on human population outcomes are of high methodological quality.
2. Health outcomes resulting from Fitbit-, Garmin-, or Apple Watch-based interventions have high certainty of evidence.

2. Materials and Methods

The protocol of this systematic review was registered in advance in PROSPERO (ID: CRD42021276533). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was followed in this study [26].

2.1. Eligibility Criteria

Full text systematic reviews originally published in English in peer-reviewed journals describing meta-analyses of all clinical outcomes of Fitbit-, Garmin-, or Apple Watch -based interventions compared with other comparators without Fitbit, Garmin or Apple Watch. In other words, studies examining the effect of Fitbit, Garmin, or Apple Watch on human populations of both genders and any age group were included. This study was not limited to a specific population, time period, clinical setting, or geography. Animal-based studies, studies of mixed interventions and the accuracy of model testing were not eligible for this study. All meta-analyses reporting clinical outcomes resulting from a Fitbit, Garmin, or Apple Watch -based intervention were included.

2.2. Search Strategy

Systematic searches were carried out in the PubMed, Scopus, and the Web of Science electronic databases from inception to August 2021. The search was also updated on 20 October 2021, and the reference lists of the final eligible studies were also reviewed to find other meta-analyses. The subsequent steps were followed: keywords Fitbit OR Garmin OR "apple watch" were searched in the title and/or abstract in PubMed and Scopus. The mentioned search formula was performed in the Topic field in the Web of Science database. In addition, controlled vocabulary MeSH in PubMed was used while performing the search. To identify systematic reviews, the NIH search strategy was used in PubMed [27]. In Scopus, to identify systematic reviews, the database search filters were used, and the document types were restricted to "review". In the Web of Science database, the document types were also restricted to "review articles".

2.3. Study Selection

All identified documents from the searches were managed and entered into the Microsoft Excel software. Duplicates were identified by the DOI number. If no DOI number was available, the titles were used to recognize the duplicates. Titles and abstracts of included studies were screened independently by two reviewers according to the following questions: is it a systematic review study? (yes/unsure; no); is it a Fitbit, Garmin, or Apple Watch study? (yes/unsure; no). Studies for which the answers to the questions were yes/unsure were eligible for the next screening stage (full text screening). In the next stage, the reviewers assessed the studies according to the following questions: is this a study published in English? (yes; no); is this a study involving humans? (yes; no); is it a meta-analysis? (yes; no); was Fitbit, Garmin, or Apple Watch used as an intervention in the study? (yes; no). Filtering the answer to all questions by "yes" identified eligible studies. In both the title/abstract and the full-text screening, disagreements were discussed by the reviewers to reach consensus. The results of the screening and the selection of eligible studies were visualized using the PRISMA flowchart [28] (Figure 1).

2.4. Data Extraction

Data extraction from the qualifying studies was performed separately by two reviewers. The reviewers discussed to reach an agreement to resolve the disparities. The following variables were extracted from the studies by the reviewers: publication year, authors' name, first author's country using the ISO 3166-1 code, study designs of included studies, population, intervention, comparator, outcome(s), the total number of population in each outcome, effect size, and confidence intervals.

2.5. Assessment of Methodological Quality

The quality of the eligible systematic reviews was methodologically and independently assessed by two reviewers using the AMSTAR-2 tool, which contains 16 different questions with the answers: yes, partly yes or no [29]. Any discrepancies in the ratings of the 16 items of the AMSTAR-2 checklist were resolved through discussion to reach consensus. The following methodology was used to report the methodological quality of the review: qualitative responses were converted into quantitative data: 1 point for questions with answers "yes", 0.5 point for questions with answers "partly yes", and 0 points for questions with answers "no".

2.6. Assessment of the Quality of Evidence

As mentioned earlier, the quality of evidence of each outcome in meta-analyses was independently assessed by two reviewers using the GRADE tool [24], which assesses the quality of synthesis of every outcome based on five domains: 1) risk of bias [30], 2) inconsistency [31], 3) imprecision [32], 4) publication bias [33] and 5) indirectness [34]. Depending on the severity of every domain, the GRADE tool proposes a 0, 1 or 2 downgrades. Therefore, the following strategies were applied [35].

Risk of Bias: if less than 75% of the studies included in a meta-analysis had a low risk of bias or if the risk of bias of the individual studies included in a meta-analysis was not reported, 1 downgrade was considered to decide on the overall risk of bias of a meta-analysis. On the other hand, no downgrading was performed if 75% or more than 75% of the included studies in a meta-analysis had a low risk of bias.

Inconsistency: to make a decision regarding inconsistency, the reported heterogeneity (I^2) was considered in each outcome. Accordingly, 1 downgrade was assigned if the calculated heterogeneity in each outcome was reported to be at least 75%. Otherwise, no downgrade was assigned. If heterogeneity was not reported, also 1 downgrade was assigned [36].

Imprecision: following the recommendations in the GRADE Handbook [24], no downgrade was assigned to any outcome if the pooled sample size was minimum 2000. 1 downgrade was assigned if the pooled sample size was less than 200. When the size of the pooled sample was between 200 and 2000, confidence intervals were considered. If confidence intervals were not narrow, 1 downgrade was assigned. Otherwise, no downgrade was considered.

Publication bias: one of the most commonly used methods to assess publication bias in studies is the funnel plot. Due to some limitations of the funnel plot method in accurately identifying the publication bias [37,38], the strategy of Trim and Fill was used [39], which recalculates the effect size by imputing missing studies from the analysis. If the recalculated effect size was changed due to some missing documents, 1 downgrade was assigned. Alternatively, no downgrade was assigned.

Indirectness: to assess indirectness in each outcome, 1 or 2 downgrades was considered if there were differences between the included studies in each outcome in terms of intervention, population, or comparator, depending on the severity of the differences. If the included studies were consistent with the review questions in each outcome and were coherent with regards to population, intervention, or comparator, no downgrading was considered [40].

Finally, the quality of every evidence was reported at four levels: 0 downgrade as a high quality, 1–2 downgrades as a moderate quality, 3–4 downgrades as a low quality, and 5–6 downgrades as a very low quality of synthesis [36].

3. Results

3.1. Literature Search

The literature search yielded 16, 83, and 41 entries in PubMed, Scopus, and the Web of Science, respectively. After removing duplicates, 80 studies were eligible for title/ab-

stract screening. Twenty-six records were excluded, of which 21 records were not considered systematic reviews, and five studies were excluded because they did not focus on Fitbit, Garmin, or Apple Watch. Consequently, 54 studies progressed to the full-text screening stage. One study because it was not in English, 42 records because they did not report a meta-analysis, and 10 screened documents because they did not include Fitbit, Garmin, or Apple Watch as an intervention did not conform to the study. Because the goal of the current evaluation was to examine the methodological and evidence synthesis quality of meta-analyses evaluating the effect of Fitbit, Garmin, or Apple Watch as an intervention on changing human health outcomes, the accuracy of model testing studies were not suitable for this study. For example, research that provided a meta-analysis of the accuracy of Fitbit models in evaluating sleep was excluded [41]. Overall, one systemic review that reported a meta-analysis of the effect of Fitbit as an intervention on population health outcomes compared with non-Fitbit, Garmin, or Apple Watch met the inclusion criteria and was included in the final analysis[41] (see the PRISMA flow diagram, Figure 1).

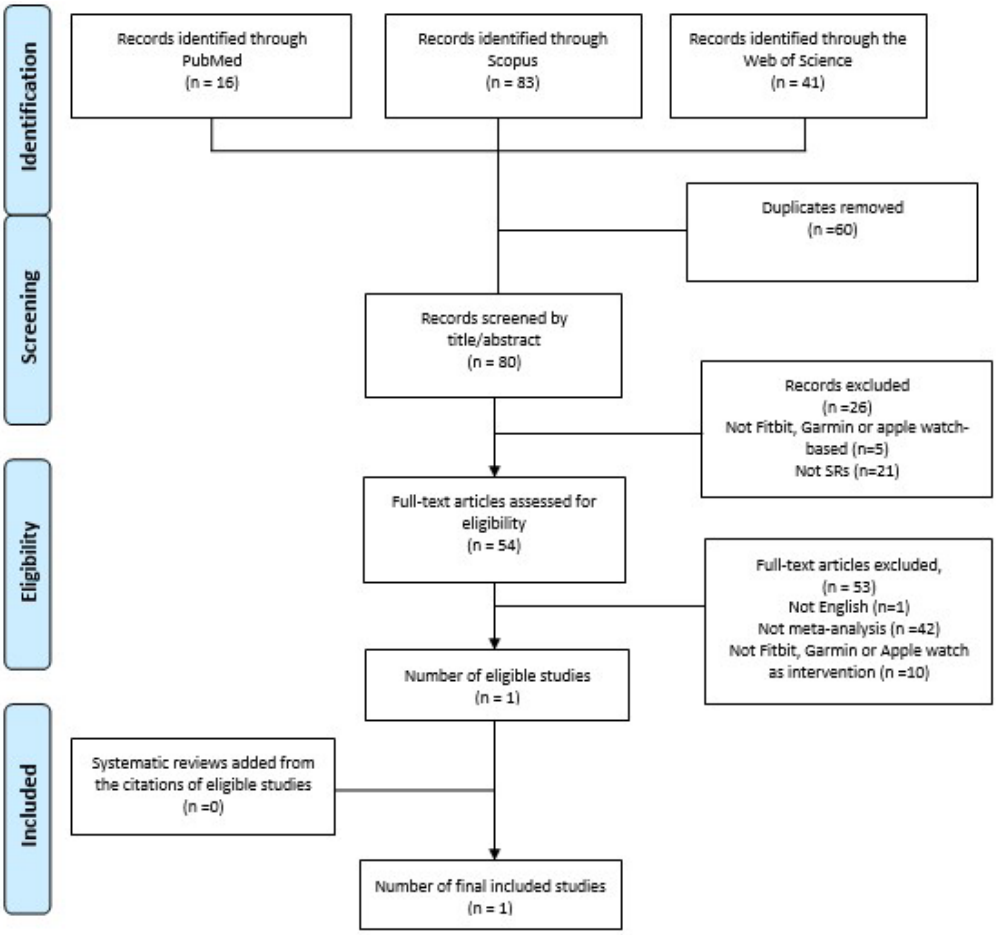

Fig. 1. PRISMA flowchart of the study selection.

3.2. Characteristics of the Included Systematic Reviews

The characteristics of the included systematic review (SR) that met the inclusion criteria are summarized in Table 1. The review was published in 2020 [42]. The number of included primary studies in the review was 41 documents. The study included healthy subjects, patients with cardiovascular risks, cardiometabolic or chronic diseases, comparing the effect of Fitbit with methods without a device.

Table 1. Characteristics of the included systematic review.

Author (year)	Country	Study design	Population	Intervention	Comparator	Outcome(s)	Conclusion summary
Ringeval M. et al. (2020)	Canada	RCT[1]	Healthy subjects, patients with cardiovascular risks, chronic diseases or cardiometabolic diseases	Fitbit	Without a device	Steps per day, physical activity (MVPA, min/day), weight (kg)	Using Fitbit as an intervention has the potential to improve a healthy lifestyle concerning physical activity and weight.

[1]: *RCT: Randomized controlled trial*

3.3. Methodological Quality Results

The Amstar-2 assessment showed that the review had two items that did not meet the criteria (overall score 13.5 out of 16). The methodological limitations arose from two items: questions seven and ten. Moreover, "partial yes" was assigned to question four. Due to the fact that there is one critical flaw [Item 7] with one non-critical weakness [Item 10], the review was rated as "low" in the methodological quality[29]. More details can be found in Table 2.

Table 2. Assessment of the methodological quality of review.

Study	Items																Overall
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	score
	Y	Y	Y	PY	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	13.5(L)

AMSTAR-2 Items:

1. Did the research questions and inclusion criteria for the review include the components of PICO?
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?
3. Did the review authors explain their selection of the study designs for inclusion in the review?
4. Did the review authors use a comprehensive literature search strategy?
5. Did the review authors perform study selection in duplicate?
6. Did the review authors perform data extraction in duplicate?
7. Did the review authors provide a list of excluded studies and justify the exclusions?
8. Did the review authors describe the included studies in adequate detail?
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?
10. Did the review authors report on the sources of funding for the studies included in the review?
11. If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
15. If they performed a quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?
17. Y: Yes. PY: partially yes. N: No. L: Low methodological quality

3.4. GRADE Results

The qualified review included five outcomes related to the effect of Fitbit on steps per day, physical activity, weight, objectively measured as well as self-reported sedentary behavior in a group of healthy subjects, patients with cardiovascular risks, chronic or cardiometabolic diseases. The quality assessment of the evidence synthesis showed that no outcome had high quality evidence (0%). Two of the outcomes (40%) had moderate-quality evidence synthesis, while the other three outcomes (60%) demonstrated low-quality evidence, as shown in Table 3.

Table 3. GRADE evaluation results.

Interventions	Outcomes	Effect size / (95% confidence intervals)	Risk of Bias	Inconsistency	Imprecision	Publication bias	Indirectness	Overall quality rating
Fitbit vs without a device	Steps per day	950.54 (MD)/ (475.89, 1425.18)	-1	0	-1	0	0	Moderate
Fitbit vs without a device	moderate to vigorous physical activity (MVPA; min/day)	6.16 (MD)/ (2.80, 9.51)	-1	0	-1	-1	0	Low
Fitbit vs without a device	difference in weight (kg)	-1.48 (MD)/ (-2.81, -0.14)	-1	0	-1	-1	0	Low
Fitbit vs without a device	sedentary behaviors (min/day)-objectively measured	-10.62 (MD)/ (-35.50, 14.27)	-1	0	-1	-1	0	Low
Fitbit vs without a device	sedentary behaviors (min/day)-self reported	-0.11 (SMD)/ (-0.48, 0.26)	-1	0	0	0	0	Moderate

With respect to the step-per-day outcome, which had a moderate quality of evidence, a total of two downgrades were assigned for risk of bias and imprecision, as at least 75% of the included studies did not have a low risk of bias in this endpoint. All included studies did not meet the criteria for blinding of participants and personnel criteria. In addition, because the total sample size in the synthesis ranged from 200 to 2000, the confidence intervals (475.89, 1425.18), which were not regarded as narrow, were considered.

The physical activity outcome was graded as low-quality evidence because there was a risk of bias in blinding participants and personnel in all included RCTs. Also, a downgrade in imprecision was assigned, as the sample size was 1073, and the confidence intervals were not narrow (2.80, 9.51). In addition, another downgrade was assigned because the recalculated effect size was altered by the trim-and-fill method, as a missing study was observed in the analysis. Regarding the weight outcome, three downgrades were assigned due to risk of bias, imprecision, and publication bias. Consequently, this endpoint provides low-quality evidence. The reasons were: no blinding of participants and personnel in the RCTs, the total number of samples (909) and wide confidence intervals (-2.81, -0.14),

imputation of two missing studies, which changed the recalculated effect size using the trim-and-fill method. In the case of objectively measured sedentary behavior, it also provides low-quality evidence based on risk of bias, imprecision and publication bias. No blinding of participants and personnel, small number of pooled samples (173), imputation of a missing study, which changed the recalculated effect size using the trim-and-fill method. According to the results, self-reported sedentary behaviors outcome provides moderate quality evidence. Only one downgrade was assigned to the risk of bias because participants and staff were not blinded in the two included RCTs. There was no inconsistency in the outcomes as all measured heterogeneity values (I^2) were less than 75%. Also, no indirectness was found in the included RCTs in any outcome.

4. Discussion

In evidence-based medicine, methodological evaluation and assessment of the quality of evidence are strongly recommended before medical decision making [43, 44]. High methodological quality systematic reviews providing high certainty are considered the most important sources providing the highest level of evidence [45, 46] and influencing medical decision making [47]. To the best of our knowledge, the current study is the first to assess the methodological and evidence quality of systematic reviews providing meta-analyses of the effect of Fitbit, Garmin, or Apple Watch on people's health outcomes which intends to reveal the methodological quality of systematic reviews to help clinicians make better clinical decisions.

In accordance with the inclusion criteria, only one study was ultimately included in the current review. Studies evaluating the clinical effect of Fitbit, Garmin, or Apple Watch as an intervention on the human's health outcomes were eligible. However, there are published systematic reviews with a qualitative data synthesis on the effects of smartwatches on human health that are not consistent with the present inclusion criteria [13, 48, 49]. In most meta-analyses, the above mentioned devices were used as an intervention in combination with other wearables [11, 12, 50–55], while a preliminary search on clinicalreial.gov revealed numerous completed studies. For example, the role of Fitbit as an intervention in knee replacement patients, hospitalized general patients, obese women, ICD patients, predicting risk of preterm birth, as an activity tracker during chemotherapy for breast cancer, etc., which signifies that scholars in this field should conduct more meta-analyses to determine the current evidence. In addition, a similar search found only one study that examined the clinical effectiveness of Garmin as a single intervention to predict outcomes in high-risk perioperative patients, suggesting that more clinical trials need to be conducted to evaluate the clinical effectiveness of Garmin on health outcomes. Four completed studies were also identified that examined the role of the Apple Watch in cardiac rhythm detection, electrocardiographic diagnostic performance, cardiac arrhythmia detection, and the effectiveness of the Apple Watch in reducing delayed or missed meal boluses.

Given the widespread adoption of wearables, especially smartwatches [2, 3], and the fact that Fitbit, Garmin and Apple Watch are recognized as 3 of the top 5 fitness trackers [56], they can play a prominent role in conducting clinical trials as an intervention to change the health status of populations [42]. However, as mentioned earlier, most meta-analyses have evaluated the clinical effect of these devices as an intervention by combination with other wearables, indicating a gap in this area, suggesting that researchers should conduct further clinical trials and meta-analyses on the effect of these devices as a single intervention on the health status change in different patient groups.

As the results revealed, the included systematic review had low methodological quality, indicating the need for quality improvement. The main reasons for this problem arise from two criteria. The listing of omitted studies and the justification for each study's exclusion is one of the most significant aspects to examine in the Amstar-2 tool. This deficiency is evident in most systematic reviews [57–62]. Because the listing excluded studies, and justifying the exclusion reasons for each study is a critical domain in the methodology

of systematic reviews according to the Amstar-2 guidelines [29], researchers are strongly encouraged to list studies with reasons for exclusion when conducting systematic reviews.

Another criterion is reporting the source of funding for studies included in systematic reviews. The eligible systematic review included in this study could not meet this criterion. This result of the research is consistent with the findings of some previous investigations [58–63]. On the other hand, the research conducted in 2020 on systematic reviews on spine surgery found that about 70% of the systematic reviews in this field could meet these criteria [57], which is inconsistent with our results.

Besides, 60% of the outcomes was found to have low-quality evidence, meaning that the actual effect could significantly differ from the assessed effect in the outcomes. Of the five main factors assessed, the risk of bias was identified as the most common factor (n = 5, 100%) reducing the quality of the evidence, which suggests that researchers should pay close attention to concealing attribution, using methods of blinding to reduce the impact of limitations on outcome indicators. In addition, an inadequate sample size resulted in imprecision in 80% of the outcomes, reducing the quality of the evidence. This means that researchers should pay attention to optimal information size (OIS) [64] when conducting meta-analyses. In addition, publication bias was found in 60% of the results, possibly due to a failure to search comprehensively [65], as the included study did not search the grey literature.

According to the results of the included systematic review, Fitbit had a significant effect on increasing steps per day (MD, 950/54, CI: 475.89, 1425.18) [42]. Nevertheless, as the data analysis of the result of the current study showed, this outcome is moderately certain, which means that the actual effect is probably close to the estimate of the effect. Still, it is also possible to be substantially different [66]. In addition, the results of the meta-analysis of the included systematic review [42] demonstrated the significant efficacy of Fitbit on moderate to vigorous physical activity (MD, 6/16, CI: 2.80, 9.51). However, it should be noted that the certainty of this evidence is low due to the result of GRADE on this outcome. In addition, the effectiveness of Fitbit on weight difference (MD, -1/48, CI: -2.81, -0.14) and objectively measured sedentary behavior (MD, -10/62, CI: -35.50, 14.27) was confirmed in the included study[42], but both had low certainty. This suggests that the actual effect may significantly differ from the effect estimates [66]. On the other hand, Fitbit had no significant effect on self-reported sedentary behavior (SMD, -0/11, CI: -0.48, 0.26) according to the results of the included review study [42]. Because the quality assessment of this evidence using GRADE revealed moderate certainty, one can be moderately confident in this effect estimate [66].

To answer the first hypothesis, it should be noted that the only included systematic review had low methodological quality according to the current study results. As for the second hypothesis, the application of the GRADE tool showed that none of the outcomes measured by Fitbit had a high quality of evidence. Accordingly, the effect of Fitbit in increasing daily steps and improving self-reported sedentary behavior has a moderate quality of evidence. Regarding moderate to vigorous physical activity, weight difference, and objectively measured sedentary behavior, the effect of Fitbit on the above outcomes had low certainty.

5. Conclusions

Overall, one systematic review was included in this study, and it was classified as being of low methodological quality by the AMSTAR-2 tool. From the GRADE classification of the results of the meta-analyses in the included SR, there were three case of low-quality evidence including moderate to vigorous physical activity, weight, and objectively measured sedentary behavior. The other two outcomes, which included steps per day and self-reported sedentary behaviors, had moderate-quality evidence. Regarding the results of meta-analyses, regardless of the significance or insignificance of the effect sizes, the degree of certainty of the evidence should be determined and considered before a decision is made. As the current study results showed, there is moderate and low certainty about

the effect of Fitbit on daily steps, moderate to vigorous physical activity, weight management, and sedentary behavior. Researchers in the field are encouraged to conduct more clinical studies and meta-analyses and should continue their education and pursue the AMSTAR-2 guidelines and GRADE to design high grade studies in the future that will definitely affect decisions about the use of activity trackers like Fitbit.

Strength of the study

Two well-validated tools were used, namely Amstar-2, a new version of AMSTAR, and GRADE, to assess the methodological quality as well as the quality of evidence synthesis of the literature, which enhances the quality of the research. Additionally, the search was conducted in three of the most reliable electronic databases, namely PubMed, Scopus, and Web of Science, and it was not restricted to a specific time period, which improved the quality of the search. Reviewing the reference lists of the final eligible studies may also be considered an additional aspect in determining the robustness of the search.

Limitations of the study

The findings of the current study should be evaluated in light of its limitations. To find similar research, the keywords Fitbit, Garmin, and Apple Watch were used. The sensitivity and specificity of the search algorithm in retrieving all relevant articles were not analyzed or verified, but these three keywords were hypothesized to be of potential use in obtaining similar articles. Moreover, the small number of studies ultimately found may be considered another limitation. According to the registered protocol, only one systematic review that evaluated the effectiveness of Fitbit across five distinct outcomes met the inclusion criteria. Using the search strategy outlined above, no studies could be found that examine the effect of Garmin or Apple Watch on health outcomes.

References

1. Fuller D, Colwell E, Low J, Orychock K, Ann Tobin M, Simango B, Buote R, van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. JMIR mHealth uHealth. 2020;8(9):1–23. DOI: 10.2196/18694
2. Costello K. Gartner says worldwide wearable device sales to grow 26 percent in 2019 [Internet]. 2018 [cited 2021 Aug 3]. Available from: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-
3. Savekar A, Kumar V. Industrial wearables market by device type (smart watches, glasses, headsets, and smart bands), component (sensors, displays, connectivity, processors, and others), and industry vertical (manufacturing, automotive, aerospace & defense, oil & gas, power [Internet]. 2021 [cited 2021 Jul 19]. Available from: https://www.alliedmarketresearch.com/industrial-wearables-market-A06872
4. Macridis S, Johnston N, Johnson S, Vallance JK. Consumer physical activity tracking device ownership and use among a population-based sample of adults. PLoS One. 2018;13(1):1–11. DOI: 10.1371/journal.pone.0189298
5. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: Results of a national cross-sectional survey. BMJ Open. 2016;6(7):1–8.
6. Statista. Adult wearable users penetration rate in the United States from 2016 to 2022 [Internet]. 2021 [cited 2021 Aug 15]. Available from: https://www.statista.com/statistics/793800/us-adult-wearable-penetration/
7. Washington V. Apple solidifies its lead as global smartwatch market grows 35 percent YoY in Q1 2021: Counterpoint [Internet]. 2021 [cited 2021 Aug 10]. Available from: https://gadgets.ndtv.com/wearables/news/apple-watch-smartwatch-market-lead-global-market-grow-35-percent-yoy-q1-2021-counterpoint-2450529
8. PARK J. Fitbit joins Google [Internet]. 2021 [cited 2021 Aug 1]. Available from: https://blog.fitbit.com/2021-update/?utm_source=go_mar&utm_medium=blog&utm_country_code=en_US&utm_campaign=web

9. Evenson KR, Spade CL. Review of Validity and Reliability of Garmin Activity Trackers. J Meas Phys Behav 2020;3(2):170–185. DOI: 10.1123/jmpb.2019-0035
10. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One. 2021;16(5 May):1–12. DOI: 10.1371/journal.pone.0251975
11. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Heal Promot. 2019;33(5):778–791. DOI: 10.1177/0890117118816278
12. Brickwood KJ, Watson G, O'brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: Systematic review and meta-analysis. JMIR mHealth uHealth. 2019;7(4). DOI: 10.2196/11819
13. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: A systematic review. JMIR mHealth uHealth. 2016;4(4):1–12. DOI: 10.2196/mhealth.6540
14. VanWormer JJ, Pronk NP, Kroeninger GJ. Clinical counseling for physical activity: Translation of a systematic review into care recommendations. Diabetes Spectr. 2009;22(1):48–55. DOI: 10.2337/diaspect.22.1.48
15. Foster C, Hillsdon M, Thorogood M, Kaur A, Wedatilake T. Interventions for promoting physical activity. Cochrane Database Syst Rev. 2005;2013(1):1–90. DOI: 10.1002/14651858.CD003180.pub2
16. Perez-Pozuelo I, Zhai B, Palotti J, Mall R, Aupetit M, Garcia-Gomez JM, Taheri S, Guan Y, Fernandez-Luque L. The future of sleep health: a data-driven revolution in sleep science and medicine. npj Digit Med Springer US. 2020;3(1):1–15. DOI: 10.1038/s41746-020-0244-4
17. Wright SP, Collier SR, Brown TS, Sandberg K. An analysis of how consumer physical activity monitors are used in biomedical research. FASEB J. 2017;31(S1):1020.24-1020.24. DOI: 10.1096/fasebj.31.1_supplement.1020.24
18. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. PLoS Med. 2016;13(2):1–9. DOI: 10.1371/journal.pmed.1001953
19. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Young Yoo J, Hamilton CB, Li LC. Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data. JMIR mHealth uHealth. 2018;6(8). DOI: 10.2196/10527
20. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. J Biomed Inform. 2016;63:269–276. DOI: 10.1016/j.jbi.2016.09.001
21. Emberson MA, Lalande A, Wang D, Mcdonough DJ, Liu W, Gao Z. Effectiveness of smartphone-based physical activity interventions on individuals' health outcomes: A systematic review. Biomed Res Int. 2021;2021. DOI: 10.1155/2021/6296896
22. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924–926. DOI: 10.1136/bmj.39489.470347.AD
23. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TTT, Vist GE, Williams JW. Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches. BMC Health Serv Res. 2004;4:1–7. DOI: 10.1186/1472-6963-4-38
24. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from https://gdt.gradepro.org/app/handbook/handbook.html
25. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane handbook for systematic reviews of interventions. Cochrane Handb Syst Rev Interv. 2019. DOI: 10.1002/9781119536604
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372. DOI: 10.1136/bmj.n71
27. NIH. Search Strategy Used to Create the PubMed Systematic Reviews Filter; 2018.
28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Med. 2009 Jul 21;6(7):e1000100. DOI: 10.1371/journal.pmed.1000100
29. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: A critical appraisal tool for systematic reviews that

include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:1–9. DOI: 10.1136/bmj.j4008

30. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence - Study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407–415. DOI: 10.1016/j.jclinepi.2010.07.017
31. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 7. Rating the quality of evidence – Inconsistency. J Clin Epidemiol. 2011;64(12):1294–1302. DOI: 10.1016/j.jclinepi.2011.03.017
32. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence - Imprecision. J Clin Epidemiol. 2011;64(12):1283–1293. DOI: 10.1016/j.jclinepi.2011.01.012
33. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines: 5. Rating the quality of evidence - Publication bias. J Clin Epidemiol. 2011;64(12):1277–1282. DOI: 10.1016/j.jclinepi.2011.01.011
34. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ. GRADE guidelines: 8. Rating the quality of evidence – Indirectness. J Clin Epidemiol. 2011;64(12):1303–1310. DOI: 10.1016/j.jclinepi.2011.04.014
35. Motahari-Nezhad H, Péntek M, Gulácsi L, Zrubka Z. Outcomes of digital biomarker-based interventions: Protocol for a systematic review of systematic reviews. JMIR Res Protoc. 2021 Nov 24;10(11):e28204. DOI: 10.2196/28204
36. Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, Van Wijck F, Wiffen PJ. An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. J Clin Epidemiol. Elsevier Ltd; 2016;70:106–110. DOI: 10.1016/j.jclinepi.2015.08.013
37. Zhang Y, Coello PA, Guyatt GH, Yepes-Nuñez JJ, Akl EA, Hazlewood G, Pardo-Hernandez H, Etxeandia-Ikobaltzeta I, Qaseem A, Williams JW, Tugwell P, Flottorp S, Chang Y, Zhang Y, Mustafa RA, Rojas MX, Xie F, Schünemann HJ. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences – inconsistency, imprecision, and other domains. J Clin Epidemiol. 2019;111:83–93. DOI: 10.1016/j.jclinepi.2018.05.011
38. Mavridis D, Salanti G. How to assess publication bias: Funnel plot, trim-and-fill method and selection models. Evid Based Ment Health. 2014;17(1):30. DOI: 10.1136/eb-2013-101699
39. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method. Biometrics. 2000;56(June):455–463. DOI: 10.1111/j.0006-341X.2000.00455.x
40. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58(9):894–901. DOI: 10.1016/j.jclinepi.2005.01.006
41. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: Systematic review and meta-analysis. J Med Internet Res. 2019;21(11). DOI: 10.2196/16273
42. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: Systematic review and meta-analysis. J Med Internet Res. 2020;22(10). DOI: 10.2196/23954
43. Manchikanti L, Benyamin R, Helm S, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 3: Systematic reviews and meta-analyses of randomized trials. Pain Physician. 2009;12(1):35–72. DOI: 10.36076/ppj.2009/12/35
44. Lunny C, Brennan SE, McDonald S, McKenzie JE. Evidence map of studies evaluating methods for conducting, interpreting and reporting overviews of systematic reviews of interventions: rationale and design. Syst Rev. 2016 Dec 6;5(1):4. DOI: 10.1186/s13643-015-0178-0
45. Ganeshkumar P, Gopalakrishnan S. Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare. J Fam Med Prim Care. 2013;2(1):9. DOI: 10.4103/2249-4863.109934
46. Johnson BT, Hennessy EA. Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. Soc Sci Med. 2019;233(March):237–251. DOI: 10.1016/j.socscimed.2019.05.035

47. Yousefi Nooraie R, Lohfeld L, Marin A, Hanneman R, Dobbins M. Informing the implementation of evidence-informed decision making interventions using a social network analysis perspective; a mixed-methods study. BMC Health Serv Res. 2017;17(1):1–14. DOI: 10.1186/s12913-017-2067-9
48. Jo A, Coronel BD, Coakes CE, Mainous AG. Is There a Benefit to Patients Using Wearable Devices Such as Fitbit or Health Apps on Mobiles? A Systematic Review. Am J Med. 2019;132(12):1394-1400.e1. DOI: 10.1016/j.amjmed.2019.06.018
49. Blount DS, McDonough DJ, Gao Z. Effect of wearable technology-based physical activity interventions on breast cancer survivors' physiological, cognitive, and emotional outcomes: A systematic review. J Clin Med. 2021;10(9). DOI: 10.3390/jcm10092015
50. Yen HY, Chiu HL. The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials. Obes Rev. 2019;20(10):1485–1493. DOI: 10.1111/obr.12909
51. Hannan AL, Harders MP, Hing W, Climstein M, Coombes JS, Furness J. Impact of wearable physical activity monitoring devices with exercise prescription or advice in the maintenance phase of cardiac rehabilitation: Systematic review and meta-analysis. BMC Sports Sci Med Rehabil. 2019;11(1):1–21. DOI: 10.1186/s13102-019-0126-8
52. Lynch C, Bird S, Lythgo N, Selva-Raj I. Changing the physical activity behavior of adults with fitness trackers: A systematic review and meta-analysis. Am J Heal Promot. 2020;34(4):418–430. DOI: 10.1177/0890117119895204
53. 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):1–13. DOI: 10.2196/15576
54. Liu JYW, Kor PPK, Chan CPY, Kwan RYC, Sze-Ki D. The effectiveness of a wearable activity tracker (WAT)-based intervention to improve physical activity levels in sedentary older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr. 2020;91(March):104211. DOI: 10.1016/j.archger.2020.104211
55. Kamei T, Kanamori T, Yamamoto Y, Edirippulige S. The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: A systematic review and meta-analysis. J Telemed Telecare. 2020. DOI: 10.1177/1357633X20937573
56. Alonzo I. Top 5 Best Fitness Trackers of 2021: Apple Watch, Garmin, Fitbit, and MORE!. 2021 [cited 2021 Sep 1]. Available from: https://www.techtimes.com/articles/256978/20210212/top-5-best-fitness-trackers-2021-apple-watch-garmin-fitbit.htm
57. Dettori JR, Skelly AC, Brodt ED. Critically low confidence in the results produced by spine surgery systematic reviews: An AMSTAR-2 evaluation from 4 spine journals. Glob Spine J. 2020;10(5):667–673. DOI: 10.1177/2192568220917926
58. Matthias K, Rissling O, Pieper D, Morche J, Nocon M, Jacobs A, Wegewitz U, Schirm J, Lorenz RC. The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: A cross-sectional study. Heliyon. 2020;6(9):e04776. DOI: 10.1016/j.heliyon.2020.e04776
59. Kim HR, Choi CH, Jo E. A methodological quality assessment of meta-analysis studies in dance therapy using AMSTAR and AMSTAR 2. Healthc. 2020;8(4). DOI: 10.3390/healthcare8040446
60. Almeida MO, Yamato TP, Parreira P do CS, Costa LOP, Kamper S, Saragiotto BT. Overall confidence in the results of systematic reviews on exercise therapy for chronic low back pain: a cross-sectional analysis using the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) 2 tool. Brazilian J Phys Ther. Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia. 2020;24(2):103–117. DOI: 10.1016/j.bjpt.2019.04.004
61. Storman M, Storman D, Jasinska KW, Swierz MJ, Bala MM. The quality of systematic reviews/meta-analyses published in the field of bariatrics: A cross-sectional systematic survey using AMSTAR 2 and ROBIS. Obes Rev. 2020;21(5):1–11. DOI: 10.1111/obr.12994
62. Sánchez de Ribera O, Trajtenberg N, Christensen LS. Evaluating the quality of meta-analytical reviews using the AMSTAR-2: A systematic review of meta-analytical reviews regarding child sexual abuse interventions. Child Abus Negl. 2020;104(October 2019):104463. DOI: 10.1016/j.chiabu.2020.104463
63. Chugh A, Patnana AK, Kumar P, Chugh VK, Khera D, Singh S. Critical analysis of methodological quality of systematic reviews and meta-analysis of antibiotics in third molar surgeries using AMSTAR 2. J Oral Biol Craniofacial Res. 2020;10(4):441–449. DOO: 10.1016/j.jobcr.2020.07.011

64. Schünemann H, Brożek J, Guyatt G, Oxman A. Introduction to GRADE Handbook Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach [Internet]. 2013 [cited 2021 Jan 15]. Available from: https://gdt.gradepro.org/app/handbook/handbook.html#h.9rdbelsnu4iy
65. Liang M, Yan L, Mei Z, Luo Y, Hou X, Feng Z. Methodological and reporting quality evaluation of meta-analyses on the Chinese herbal preparation Zheng Qing Feng Tong Ning for the treatment of rheumatoid arthritis. BMC Complement Med Ther. 2020;20(1):195. DOI: 10.1186/s12906-020-02978-5
66. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64(4):401–406. DOI: 10.1016/j.jclinepi.2010.07.015

Author Contributions: Study Design, HM-N; Data Collection, HM-N, AS; Statistical Analysis, HM-N; Data Interpretation, HM-N, FZ, HA; Manuscript Preparation, HM-N; Literature Search, HM-N, AS. All authors have read and agreed to the published version of the manuscript.

Acknowledgements: None.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.