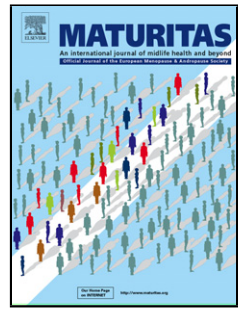
Accepted Manuscript



Title: The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review

Authors: Nicola Straiton, Muaddi Alharbi, Adrian Bauman, Lis Neubeck, Janice Gullick, Ravinay Bhindi, Robyn Gallagher

PII: S0378-5122(18)30182-8
DOI: https://doi.org/10.1016/j.maturitas.2018.03.016
Reference: MAT 6985

To appear in: *Maturitas*

Received date: 9-3-2018
Revised date: 27-3-2018
Accepted date: 29-3-2018

Please cite this article as: Straiton Nicola, Alharbi Muaddi, Bauman Adrian, Neubeck Lis, Gullick Janice, Bhindi Ravinay, Gallagher Robyn.The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review.*Maturitas* https://doi.org/10.1016/j.maturitas.2018.03.016

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review

Nicola Straiton[a, b*], Muaddi Alharbi[a], Adrian Bauman[c], Lis Neubeck[d], Janice Gullick[a], Ravinay Bhindi[b,c], Robyn Gallagher[a]

a. Sydney Nursing School, University of Sydney, Sydney, NSW, Australia

b. Department of Cardiology, Royal North Shore Hospital, Sydney, NSW, Australia

c. The University of Sydney, Sydney, NSW, Australia

d. School of Health and Social Care, Edinburgh Napier University, Edinburgh, UK

***Corresponding Author**

Nicola Straiton

Address: Charles Perkins Centre, Sydney Nursing School

The University of Sydney

Room 2210, Level 2, Building D17,

The University of Sydney, NSW 2006

Tel: +61 2 9463 2402; Fax: +61 2 9463 2053

E-Mail Address: nstr4913@uni.sydney.edu.au

This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Highlights

- This is the first systematic review to explore the validity and reliability of consumer-grade activity trackers for recording step count and activity duration in older, community-dwelling adults.
- Consumer wearables are valid in the measurement of step count and duration of physical activity, as confirmed by reference monitors or gold-standard validation techniques.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- The majority of consumer wearables overestimated step count, and to a lesser extent duration of physical activity.
- Slower walking speed and impaired ambulation reduced the level of agreement between consumer wearables and reference devices.

Abstract

Objective: To understand the validity and reliability of consumer-grade activity trackers (consumer wearables) in older, community-dwelling adults.

Methods: A systematic review of studies involving adults aged over 65 years who underwent physical activity monitoring with consumer wearables. A total of 7 observational studies qualified, identified from electronic databases: MEDLINE, EMBASE, Cochrane Library and others (2014 to 2018). Validity was interpreted using correlation coefficients (CC) and percentage error for agreement between reference devices or gold-standard validation methods. Reliability was compared using mean differences or ranges (under- or overestimation) of step count and activity time.

Results: Total sample size was 290 adults, mean age of 70.2±4.8 years and females constituting 46.7±26.1%. The studies evaluated eight different consumer wearables used by community-dwelling adults with a range of co-morbidities. Daily step count for all consumer wearables correlated highly with validation criterion, especially the ActiGraph device: intraclass correlation coefficients (ICC) were 0.94 for Fitbit One, 0.94 for Zip, 0.86 for Charge HR and 0.96 for Misfit Shine. Slower walking pace and impaired ambulation reduced the levels of agreement. Daily step count captured by Fitbit Zip was on average 7117 (±5,880.6), which was overestimated by five of the eight consumer wearables compared with reference devices (range 167.6 to 2,690.3 steps/day). Measurement of activity duration was accurate compared with reference devices, yet less so than step count.

Conclusion: In older, community-dwelling adults, consumer wearables accurately measure step count and activity duration, as confirmed by reference devices and validation methods. Further

ACCEPTED MANUSCRIPT

research is required to understand how co-morbidities, gait and activity levels interact with monitoring in free-living environments.

Keywords, older, physical activity, exercise, tracker, wearable, measure.

1. Introduction

Healthy ageing has evolved from simply a desire to increase life expectancy, to more aspirational aims of avoiding disease, preserving physical functioning and allowing an independent engagement with life. Compelling evidence from meta-analyses and Cochrane reviews demonstrate the benefits of exercise, especially when physical activity (PA) is planned, structured and underpinned by the goal to improve or maintain physical fitness, performance and health for elderly people [1]. Exercise benefits for older people not only include prevention and treatment of increasingly prevalent chronic conditions, such as cardiovascular disease, but often more importantly, improvements in health-related quality of life [2]. The World Health Organisation [3] recommends a minimum average daily activity of 30 minutes at moderate-intensity, which can be achieved through walking (steps) and walking rapidly or uphill (steps/time). Free-living PA can be defined as a person's everyday physical activity in their usual environment, and for older, community-dwelling adults it is free-living PA which is crucial. The aims of PA in older people are to reduce sedentary behaviour, increase autonomy in daily activities and sustain long-term exercise goals, and as such these objectives are recommended to be achieved in older people’s normal surroundings [4].

Contemporary technology provides an unprecedented opportunity for the use of consumer-grade activity trackers (consumer wearables) to both understand, investigate and promote sustainable PA in older people. The measurement of PA is essential to all of these aims. The mainstay of measurement in older people has been self-report, and less often through the use of research-grade

ACCEPTED MANUSCRIPT

activity trackers. Self-report has inherent limitations, including the need to monitor duration and type of PA is being undertaken, as well as the capacity to recall specific aspects [5]. Often both these necessary attributes are flawed, leading to overestimation. Furthermore, self-report questionnaires have been criticised for the tendency to be age-or disease biased, excluding common elements of regular physical activity (e.g. personal care or domestic tasks), both potentially creating restrictions in accurate responses from older people. In contrast, research-grade motion sensors, which monitor activity such as pedometers, actometers and accelerometers, circumvent these issues by direct tracking [6]. Whilst these measurement devices are validated, they are often cumbersome and difficult to apply, therefore less useful in long-term monitoring of everyday PA in older community-dwelling adults, for either research requirements or personal motivation. Consequently, within the array of trackers, it is the consumer-grade physical activity trackers (e.g. Fitbit™, Polar™, Garmin™, Apple Watch Sport™), which may become the preferred self-monitoring, measurement option. Many studies to date have sought to validate the growing range of consumer-grade activity trackers, in both 'controlled' laboratory and 'free-living' environments [7,8], and with healthy and disease-specific cohorts. Evenson et al (2015) [9] published the most recent systematic review on the validity of consumer wearables to monitor PA, and reported at the time of the search in 2014 only one study [4] had an older adult sample. Therefore, this systematic review aims to provide an update on prior evidence, with a specific focus on studies reporting the validity and reliability of consumer-grade activity trackers in older, community-dwelling adults.

2. Methods

2.1 Search Strategy

The review was guided by Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) [10], with inclusion criteria and methods of analyses decided in advance. We performed a systematic search of three electronic databases MEDLINE, CINAHL and COCHRANE Central Register

ACCEPTED MANUSCRIPT

of Controlled Clinical Trials, from November 2014 to January 2018. For the purpose of this review, the keywords searched included 'older', 'physical activity', 'exercise', 'steps', 'tracker', 'wearable', 'consumer-grade' and 'measure' amongst others. Additional manual searches of reference lists from eligible papers, related trial bibliographies, conference abstracts and Google Scholar (Figure 1) were undertaken. The eligibility of articles was independently assessed by two authors (NS, RG), and a third author (MA) made the decision when uncertainty occurred. Studies fulfilling the following criteria were included: (1) samples of community-dwelling adults (2) age >65 years, or having a sample mean age of >65 years, (3) measurement of step count with or without activity duration (time), using at least one consumer-grade activity tracker (wearable) (excluding pedometer only devices) and (4) the monitored period was in a 'free-living' or 'controlled' environment. Studies were excluded if the full-text was not available or published in a language other than English.

2.2 Data Extraction and Quality Assessment

Data was collected from eligible studies including, but not limited to, first author, publication year, location (country), sample size, age, proportion of female, chronic medical condition(s), consumer-grade tracker used, placement, step count, activity duration (minutes or hours), monitoring period, reference measure, validity and reliability outcomes. The Critical Appraisal Skills Programme CASP (Cohort Study) Checklist was used to assess quality. On assessment, 75% of the studies met a minimum 80% of the evaluation criteria and therefore no paper was excluded on the basis of this quality assessment.

2.3 Data Analysis

For any extracted information that was missing from the publication, we made an attempt to contact at least one author to obtain the information. We tabulated our data, highlighting study

ACCEPTED MANUSCRIPT

characteristics, consumer-grade activity tracker specifics, validity measures and comparative reliability of devices. Validity is reported on the basis of correlation coefficients (CC) and percentage error for agreement between devices or gold standard direct observation methodology (e.g. visual count). Reliability is reported as the mean differences or range (under or overestimation) of step count and activity duration versus similar validity comparators. In addition, we report activity monitor placement and the potential effects this had upon reliability of both devices.

3. Results

3.1 General characteristics of studies

The initial search identified 976 abstracts published since 2014, of which 956 were excluded; 20 papers had full-text assessment (Figure 1). Further examination excluded 13 studies and the remaining 7 studies were included in the review. The total sample includes 290 participants with a mean age of 70.2±4.8 years and females constitute 46.7±26.1% (Table 1). Studies were performed in Europe [14,16], Australia [11,13], UK [17], US [15] and Canada [12]. Of the samples, one recruited participants during hospitalisation yet the primary monitoring period was measured in a free-living environment [16], the remainder were community-dwelling adults. Two studies [13,16] had samples diagnosed with coronary heart disease (CHD) and one [14] had sample of chronic obstructive pulmonary disease (COPD) patients. The remainder of the studies did not have specific disease-based criteria.

3.2 Activity trackers and monitoring trends

In total, eight different consumer-grade activity trackers were included in this review (Table 2), including Fitbits One™, Zip™, Flex™ and Charge HR™, Jawbone UP™, Misfit Shine™, Omron HJ-112™ and Polar A300™. Devices were used in both combination [11,15,17] and alone [12,13,14,16].

ACCEPTED MANUSCRIPT

Validity was assessed primarily against research-grade devices including ActiGraph™, Bodymedia Sensewear™, Shimmer3™ and NL2000i™, however two studies [11,12] also used direct visual count as a comparator. Consumer-grade activity trackers were worn on the waist (50%), wrist (40%) and ankle (10%). Only two studies reported activity as overall duration [13,14], though seven of the eight consumer-grade trackers had the capability to do so. Five studies [11,13,14,16,17] required participants to monitor activity in their own free-living environments for a time period ranging from 3 to 28.2 days from placement. In two studies activity was monitored in controlled settings using pre-determined walking tracks and regulated gym conditions [12,15].

3.3 *Validity of consumer-grade activity trackers (step count and activity time)*

There was a high correlation in daily step count between the eight consumer-grade activity trackers and the reference research-grade trackers or study comparator (Table 3), over a monitoring period ≥ 24 hours in a free-living environment, in all but one study [16]. Validity was high versus ActiGraph™, the mostly commonly used reference device, with intraclass correlation coefficients (ICC) for Fitbit One ™ (0.94), Zip ™ (0.94), Charge HR ™ (0.86) and Misfit Shine™ (0.96). One study [12] that evaluated monitoring accuracy within controlled environments, reported that during speed trials when compared to a visual count, the percentage error varied depended on placement of the Fitbit One ™. At ankle level, agreement was <10% at speeds of 0.4–0.9 m/s and at waist <10% for only the 2 fastest speeds 0.8 and 0.9 m/s [12]. The conclusion was that Fitbit One ™ accurately captures steps at slow speeds when placed at the ankle. Another study [15] compared the Stepwatch ™ with four consumer-grade activity trackers worn at the same time, and reported that three (Omron HJ-112 ™, Fitbit One ™, and Jawbone UP ™) were accurate at measuring steps, in both non-impaired and impaired ambulation, older adults. Daily activity duration was also accurately captured by consumer-grade devices compared to reference trackers, although at a lower correlation than for step counts

ACCEPTED MANUSCRIPT

(Fitbit Flex™ vs ActiGraph™ MVPA (r=0.74 Pearson correlation coefficient) [13] and Polar A300™ vs Bodymedia Sensewear™ (r=0.25 ICC) [14].

3.4 Reliability of consumer-grade activity trackers (step count and activity time)

Daily step count was reported to be overestimated in the majority of cases, and regardless of device location (Table 4). Participants (mean age 67±10.03) monitored over a four-week period with a consumer-grade activity tracker walked an average 7,117 steps per day (±5,880.6) [16]. This is comparable to average daily steps captured by the ActiGraph ™ (research-grade) device in community-dwelling seniors, enrolled without a pre-requisite clinical condition (7,503.7±3,526.3 steps per day), over a seven-day monitoring period. Compared to research-grade monitors, four studies [11,13,14,17] reported an overestimation in daily step counts in five of the eight consumer-grade activity trackers (Fitbit One™, Fitbit Zip™, Fitbit Flex™, Fitbit Charge HR™ and Polar A300™), with discrepancies ranging from 167.6 to 2,690.3 [17] steps per day. In contrast, two devices underestimated daily steps taken, Fitbit Zip™ (waist by 2402.7) and Misfit Shine™ (waist by 633.2) versus research-grade devices [16,17]. Lastly, compared to research-grade monitors, physical activity duration was overestimated [13] by the Fitbit Flex™ (average 10 min/day) and highly variable when measured with the Polar A300™ device [14].

4. Discussion

Our paper which includes an additional seven studies to the most recent systematic review [9], provides further evidence contributing to the use of consumer wearables for the measurement of physical activity. To our knowledge this is the first systematic review to explore the validity and reliability of consumer-grade activity trackers for measuring step count and duration of PA in older, community-dwelling adults. Consumer wearables proved valid in the measurement of step count

ACCEPTED MANUSCRIPT

and duration of PA when compared to research-grade activity trackers or visual (step-count) reference techniques. Slower walking speed and impaired ambulation reduced agreement. Most consumer-grade activity trackers overestimated step count, and to a lesser extent duration of PA. Furthermore, in older people who have very slow walking speeds, careful consideration needs to be given to tracker selection. These are important finding given the growth in the use of consumer-grade activity trackers to monitor and improve physical activity in health-care research and for independent use by consumers.

4.1 Validity and reliability of consumer-grade activity trackers in older adults

Tracking of free-living activity allows shared decision-making amongst patients and healthcare practitioners, when evaluating intervention need versus expected quality of life gains.

With the majority of consumer-grade activity trackers evaluated effectively capturing PA in older adults, future use may serve several research and interventional purposes requiring this accurate measurement. Such devices may be substituted for research-grade devices when evaluating interventions, prescribing PA and for self-monitoring by older people to manage their own activity levels. Physical functioning is imperative to allow seniors to engage fully and independently with their environment, especially as their complex care demands change. As such, consumer wearables may have particular benefit for older adults [18].

As people age, many walk at a slower pace, experience uneven gait patterns and often require walking aids. It was at these reduced walking speeds or lower activity levels, that several studies [12,13,16] in this review reported greater percentage error or data acquisition difficulties, compared to validation devices and methods. Unfortunately, many individuals with chronic diseases, often experience common disease symptoms, such as breathlessness or fatigue (i.e. COPD or heart failure patients) and thus are prone to experience daily physical limitations [19]. In terms of device

ACCEPTED MANUSCRIPT

placement, again there was disparity across trackers. For step count, compared to the wrist location, waist positioning appears to have less misrepresentation [16,17]. Whereas for COPD patients [14], activity duration monitoring varied significantly between the wrist-worn consumer wearable and the upper arm-worn validation monitor, potentially highlighting the sensitivity of the device components. Regardless of the variability, continuous wearing of both consumer or research-grade activity trackers, at the same location, provided the most accurate step count and activity duration results.

Nonetheless, it may be these individual level variables that may pose the greatest challenges ahead. For instance, to date consumer wearables have been driven by the commercial development of fitness trackers, meaning that market guiding forces may limit capacity to influence development for use amongst older people [20]. Many consumer wearable algorithms are limited to recommended daily step counts or activity levels for that of healthy adults [21], potentially explaining the several measurement discrepancies observed in this review. In addition, as the use of consumer wearables grows, the complicated challenge of personal data management and security will become more apparent for patients and healthcare providers alike, an important yet unresolved matter. Manufacturers will need to address these issues which underpin the expansion of consumer wearable populations.

Despite high accuracy of consumer-grade activity trackers compared to reference devices, reliability, was highly variable across devices in relation to agreement of steps taken per day. If we accept that the average daily step count for a community-dwelling older adult is around 7,500 steps/day [17], over-or underrepresentation varied by as much as 30% in two separate studies [16,17] between consumer and research-grade devices. Such variation means that the results from this study are difficult to apply to the wider, community-dwelling older adult population. Yet, it is important to acknowledge the physical activity data obtained by these consumer wearables was objective, and

ACCEPTED MANUSCRIPT

recent studies have shown on average only 20% of adults aged 65-74 achieve 30mins of moderate-intensity activity at least 5 days per week [22]. Therefore, such devices may have strong potential for older people to understand their own physical capabilities and increase self-efficacy towards personal and health-related goal achievement.

4.2 Older adults, comorbidities and physical activity

One of the important strategies to reduce aging-related morbidity is to increase physical activity among older people (Bauman et al. 2016) [23]. It was therefore encouraging to see that the overall mean age (70.2±4.8) of the sample in this review was sufficiently reflective of the aging population. Also, three included studies [13,14,16] deliberately enrolled participants with chronic illnesses (cardiac and respiratory). This is important as it is within these groups of older adults that the greatest gains can be made from improved physical activity [24]. Yet despite the increasing number of studies on older adults in this review compared to the last [9], the overall sample size was less than half, highlighting a gap in research being undertaken in this field for older adults. This underrepresentation means that current guidelines are often unable to provide comprehensive evidence-based information on the treatment of many chronic conditions [25].

5. Limitations

This review has several limitations. The literature on consumer-grade activity trackers is rapidly growing and it is therefore possible that studies were missed despite our best endeavors. Also, most studies had small samples and few had sample size calculations. The participants included in this review also represent a range of medical conditions, from acute clinical events to multiple chronic, long-term illnesses. How these different variables impact directly on step counts and physical activity time is not yet possible to predict. Several included studies reported issues with missing or

ACCEPTED MANUSCRIPT

uninterpretable data across comparator devices and explanations from study investigators were not consistent, including altered positioning, low-intensity activity, loss of device and set-up error. Importantly, the commercial imperative to improve updated marketable brands of wearable devices, means that several of the devices reported may no longer be in production and more recently launched devices may become redundant before testing is complete. Despite the limitations, this review is needed because of the increasing challenge for healthcare researchers to understand and measure physical activity across the rising number of older adults in their community-dwelling environments.

6. Conclusion

This systematic review included seven studies assessing validity and reliability of consumer-grade activity trackers for measuring physical activity through step counts and activity time in older, community-dwelling adults. Overall, the consumer-grade activity trackers were highly accurate for measuring average daily step count, and to a lesser extent, actual activity duration compared to research-grade reference devices. The review also highlighted issues that need to be addressed should the use consumer-grade activity trackers continue to grow, especially in relation to slow walking speeds, device positioning and gait. Nonetheless, this review provides supporting evidence and a more nuanced understanding for the use of consumer-grade activity trackers in the assessment of functional capacity in older adults.

Contributors

Nicola Straiton conceived, designed and performed the review, analysed the data and drafted the manuscript.

Muaddi Alharbi conceived, designed and performed the review, analysed the data and drafted the manuscript.

ACCEPTED MANUSCRIPT

Adrian Bauman conceived and designed the review, analysed the data and drafted the manuscript.

Lis Neubeck conceived and designed the review, analysed the data and drafted the manuscript.

Janice Gullick contributed to data collection and analysis.

Ravinay Bhindi contributed to data collection and analysis.

Robyn Gallagher conceived, designed and performed the review, analysed the data and drafted the manuscript.

Conflict of interest

The authors report no relationships that could be construed as a conflict of interest.

Funding

No funding was received for the preparation of this review.

Provenance and peer review

This article has undergone peer review.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Boden, W.E., Franklin, B.A., Wenger, N.K. Physical activity and structured exercise for patients with stable ischemic heart disease. JAMA. 2013;309:143–144.
2. de Labra C, Guimaraes-Pinheiro C., Maseda A., Lorenzo, T., et al. Effects of physical exercise interventions in frail older adults: A systematic review of randomized controlled trials. BMC Geriatrics. 2015;15:154.
3. Marsh, A. P., Ip, E. H., Barnard, R. T. et al. Using video animation to assess mobility in older adults. The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences. 2011;66:217-227.
4. Stahl, S.T., Insana, S.P. Caloric expenditure assessment among older adults: Criterion validity of a novel accelerometry device. J Health Psych. 2014; 19(11):1382–7.
5. Cerin, E., Cain, K.L., Oyeyemi, A.L., et al. Correlates of Agreement between Accelerometry and Self-reported Physical Activity. Medicine and science in sports and exercise. 2016;48(6):1075-1084.
6. Najafi, B., Armstrong, D.G., Mohler, J. Novel wearable technology for assessing spontaneous daily physical activity and risk of falling in older adults with diabetes. J Diabetes Sci Technol 2013;7:1147-1160.
7. Fulk, G.D., Combs, S.A., Danks, K.A., et al. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. Physical Therapy. 2014;94(2), 222–229.
8. Kooiman, T.J.M., Dontje, M.L., Sprenger, S.R., et al. Reliability and validity of ten consumer activity trackers. BMC Sports Science Medicine and Rehabilitation. 2015;7:24.
9. Evenson, K.R., Goto, M.M., Furberg, R.D. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. 2015;12(1):1–22.
10. Liberati, A., Altman, D.G., Tetzlaff, J., et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62:e1–34.

ACCEPTED MANUSCRIPT

11. Paul, S.S., Tiedemann, A., Hassett, L.M. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport Exerc Med 2015;1: 1–5.
12. Simpson, L.A., Eng, J.J., Klassen, T.D., Lim, S.B., et al. Capturing step counts at slow walking speeds in older adults: Comparison of ankle and waist placement of measuring device. J Rehabil Med. 2015;47(9):830–5.
13. Alharbi, M., Bauman, A., Neubeck, L., et al. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016;23:1476–85.
14. Boeselt, T., Spielmanns, M., Nel,l C., et al. Validity and usability of an activity monitor for monitoring physical activity in patients with chronic obstructive pulmonary disease (COPD). 2016;11(6): e0157229.
15. Floegel, T.A. et al. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. J. Gerontol. Ser. A: Biol. Sci. Med. Sci. 2016;72(2):229-236.
16. Thorup, C., Andreasen, J., Sørensen, E.E.,et al. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. BMJ Open. 2017;7(3):e011742.
17. Farina, N., Lowry, R. G. The Validity of Consumer-Level Activity Monitors in Healthy Older Adults in Free-Living Conditions. J Aging Phys Act. 2018;26(1):128-135.
18. Godfrey, A. Wearables for independent living in older adults: Gait and falls. Maturitas 2017;100:16–26.
19. Waschki, B., Kirsten, A. M., Holz, O. Disease Progression and Changes in Physical Activity in Patients with Chronic Obstructive Pulmonary Disease. BMC Med. 2015;13:77
20. Chiauzzi, E., Rodarte, C., DasMahapatra, P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Medicine. 2015;13:77.

ACCEPTED MANUSCRIPT

21. Piwek, L., Ellis, D.A., Andrews, S., Joinson, A. The Rise of Consumer Health Wearables: Promises and Barriers. PLoS Medicine. 2016;13(2):e1001953.

22. French, D. P., Olander, E. K., Chisholm, A., et al. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Annals of Behavioral Medicine, 2014;48(2):225–234.

23. Bauman, A., Merom, D., Bull, F.C. et al. Updating the evidence for physical activity: Summative reviews of the epidemiological evidence, prevalence, and interventions to promote "active aging". Gerontologist. 2016;56:S268–S280.

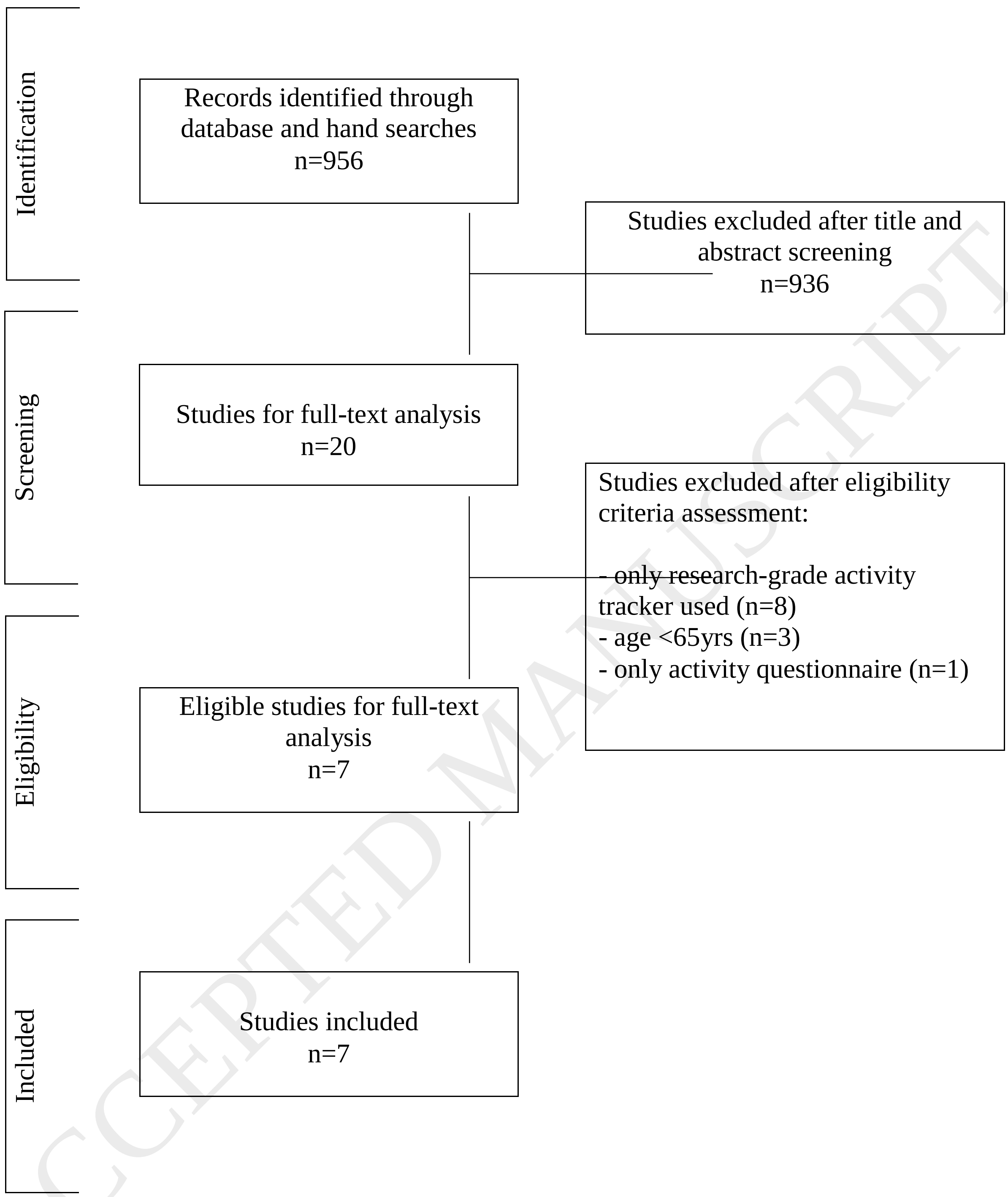
24. O'Gara, P.T., Kushner, F.G., Ascheim, D.D., et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2013;61:e78–e140.

25. Rich, M.W., Chyun, D.A., Skolnick, A.H., et al. American Heart Association Older Populations Committee of the Council on Clinical Cardiology, Council on Cardiovascular and Stroke Nursing, Council on Cardiovascular Surgery and Anesthesia, and Stroke Council; American College of Cardiology; and American Geriatrics Society. Knowledge gaps in cardiovascular care of the older adult population: a scientific statement from the American Heart Association, American College of Cardiology, and American Geriatrics Society. Circulation. 2016;133(21):2103-2122.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Fig 1. Search Strategy



ACCEPTED MANUSCRIPT

Table 1. Study Characteristics

First author, year, country	Sample size (n)	Mean age (yrs), ± SD	Female (%)	Consumer-grade activity tracker(s)	Monitoring setting	Monitoring period (days)	Inclusion criteria	Medical condition(s)
Paul, S. S. et al (2015), Australia [11]	32	67.7±5.7	63.0	Fitbit One Fitbit Zip	free-living	7	Participants were 1) aged 60+ years, 2) living at home, 3) regular (weekly) users of the internet via a computer or tablet device and 4) regular (at least once/week) out of home activity without physical assistance from another person.	41% reported 1-2 comorbidities (diagnoses not reported)
Simpson, L. A. et al. (2015), Canada [12]	42	73.0±6.9	74.0	Fitbit One	controlled	<1	Participants were 1) aged 65+ years and 2) able to walk independently for at least 30 m with/without an assistive device.	26% reported diagnosed medical conditions (most often: high blood pressure, diabetes, and thyroid disorders)
Alharbi, M. et al. (2016), Australia [13]	48	65.6±6.9	47.9	Fitbit Flex	free-living	4	Participants were 1) diagnosed CHD or family 2) completed phase II cardiac rehabilitation; 3) engaged in regular physical activity of 30 minutes/day; 4) able to participate for the full four-day study and 5) able to apply and wear both devices simultaneously.	CHD
Boeselt, T. et al. (2016), Germany [14]	20	66.4±7.4	15.0	Polar A300	free-living	3	Participants were 1) were aged between 40 to 90 years, 2) diagnosed with COPD stage I to IV and 3) able to walk.	COPD
Floegel, T. A. et al. (2017), US [15]	99	78.9±8.6	71.0	Fitbit One, Fitbit Flex, Omron HJ-	controlled	<1	Participants were 1) aged 62+ years, 2) met the health conditions on modified Physical Activity Readiness Questionnaire-Revised including stable	not reported

ACCEPTED MANUSCRIPT

				112, Jawbone UP			blood pressure for 3 months, no chest pain with activity or joint problems limiting physical activity, 3) able to walk 100 m without stopping with/without an assistive device and 4) willing to wear all activity monitors.	
Thorup, C. B. et al (2017), Denmark [16]	24*	67±10.03	8.3	Fitbit Zip	controlled *and* free-living	1 (post-surgery) *and* mean 28.2 (range 26–31)	Participants were 1) age ≥18 years and 2) hospitalised with acute coronary syndrome, heart failure (ejection fraction <40%), coronary artery bypass grafting or valve surgery.	CHD and valvular heart disease
Farina, N. et al (2018), UK [17]	25	72.5±4.9	48.0	Misfit Shine, Fitbit Charge HR	free-living	7	Participants were 1) aged 65-84 years, 2) community-dwelling and 3) independently ambulatory (walking aids excluded).	Charlson co-morbidity 11.7 score

* older participants (not younger healthy controls), SD = standard deviation *coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD)

ACCEPTED MANUSCRIPT

Table 2. Consumer-grade activity trackers device characteristics

Tracker	Selected Measurement	Placement	Size (cm)	Weight (g)	Cost AUD ($)
Fitbit One	Steps, distance, calories, active minutes, sleep, altimeter	Waist, pocket, bra	4.8(h) × 1.9(w) × 1.0(d)	9	129.95
Fitbit Charge HR	Steps, distance, calories, active minutes, sleep	Wrist	Large: 16.1–19.4(c) × 2.1(w) Extra Large: 19.4–23.0(c) × 2.1(w)	23	249.95
Fitbit Flex	Steps, distance, calories, active minutes, sleep, altimeter, heart rate	Wrist	Small: 14.0–17.6(c) × 1.4(w) Large: 16.1–20.9(c) × 1.4(w)	13 15	94.00
Fibit Zip	Steps, distance, calories, active minutes	Waist, pocket, bra	3.6(h) × 2.9(w) × 1.0(d)	8	79.95
Jawbone UP	Steps, distance (app), calories, sleep	Wrist	Small: 14.0–15.5 Medium: 15.5–18.0 Large: 18.0–20.0	19 21 23	*Discontinued December 2011.*
Misfit Shine	Steps, distance, calories, active minutes, sleep	Wrist, waist, necklace, bra	2.8(h) x 2.8(w) x 0.3(d)	9	89.95
Omron HJ-112	Steps, distance, calories, active minutes	Waist	1.5(h) x 5.3(w) x 3.6(d)	82	110.72

!

ACCEPTED MANUSCRIPT

Table 3. Validity Outcomes

First author, year, location	Criterion measure(s)	Consumer-grade activity tracker(s)	Consumer-grade activity tracker(s) placement	Objective measures	Validity analysis	Outcomes
Paul, S. S. et al (2015), Australia [11]	ActiGraph & visual step count (2MWT)	Fitbit One, Fitbit Zip	Waist	steps	Intraclass correlation coefficients (ICC)	Good agreement between Fitbit & ActiGraph (ICC2,1=0.66, 95% CI 0.41 to 0.82). Excellent agreement between Fitbit & ActiGraph in average steps/day over 7 days (ICC2,1=0.94, 95% CI 0.88 to 0.97). Excellent agreement between Fitbit & visually counted steps (intraclass correlation coefficient (ICC2,1) =0.88, 95% CI 0.76 to 0.94) on 2MWT.
					Percentage Error	Percentage agreement greatest for Fitbit steps vs visual count (mean 0%, SD 4%) and least for Fitbit average steps/day vs ActiGraph (mean 13%, SD 25%)
Simpson, L. A. et al. (2015), Canada [12]	visual step count (video recording)	Fitbit One	waist, ankle	steps	Percentage Error	Percentage error of Fitbit ankle < 10% at speeds of 0.4–0.9 m/s; for Fitbit waist percentage error < 10% at only the 2 fastest speeds (0.8 and 0.9 m/s)
					Mean differences	Fitbit ankle did not record zero steps at any speed. Fitbit waist recorded 0 steps for numerous participants at speeds of 0.3–0.5 m/s.
					Limits of agreement	Limits of agreement for Fitbit – Ankle narrower than limits of agreement for Fitbit –Waist at all speeds.
Alharbi, M. et al. (2016), Australia [13]	ActiGraph	Fitbit Flex	wrist	steps, MVPA	Pearson correlation	Significant correlation Fitbit Flex vs ActiGraph in males, females, total participants and cardiac patients for step counts (r 1⁄4.96; r 1⁄4.95; r 1⁄4.95; r 1⁄4.95). Less for MVPA (r 1⁄4.81; r 1⁄4.65, r 1⁄4.74; r 1⁄4.71)

					Absolute differences	Fitbit Flex over-estimated step counts in females (556 steps/day), males (1462 steps/day) and total participants (1038 steps/day). MVPA similar in females (4 min/day), males (15 min/day) and total participants (10 min/day)
					Sensitivity, specificity, PPV and AUC	Fitbit Flex high sensitivity (100% accuracy) for identifying participants achieving PA guidelines of 7000 and 10,000 steps/day cut-off points and 150min/week cut-off points. Specificity and PPV of Fitbit-Flex for these cut-off points varied (0.83-0.57 and 0.67 respectively). Fitbit Flex high accuracy and specificity for participants who failed to achieve the cut-off values above but moderate specificity and high accuracy for participants who failed to achieve 7000 steps/day.
Boeselt, T. et al. (2016), Germany [14]	Bodymedia Sensewear (SWA)	Polar A300	wrist	steps, calories, daily activity time (hrs) and METS	Intraclass correlation coefficients (ICC)	High correlations for both devices for sensed step count (r = 0.96; p < 0.01) and calories burned (r = 0.74; p < 0.01), lower correlation to daily activity (r = 0.25; p < 0.01)
					Limits of agreement	3 day data analysis showed 90% of steps (95% CI -4223 to 1887), 100% of calories (95% CI -2798 to 1887), 90% of daily activity data (95% CI -12.32, 4065) and 95% of the MET (95% CI -3.11 to 2.75) within limits of agreement
Floegel, T. A. et al. (2017), US [15]	StepWatch	Fitbit One, Fitbit Flex, Omron HJ-112, Jawbone UP	wrist, waist	steps	Intraclass correlation coefficients (ICC)	Non-impaired adults' steps underestimated by 4.4% StepWatch (ICC = 0.87), 2.6% Fitbit One (ICC = 0.80), 4.5% Omron HJ-112 (ICC = 0.72), 26.9% Fitbit Flex (ICC = 0.15), and 2.9% Jawbone UP (ICC = 0.55). Impaired adults' steps underestimated by 3.5% for StepWatch (ICC = 0.91), 1.7% Fitbit One (ICC = 0.96), 3.2% Omron HJ-112 (ICC = 0.89), 16.3% Fitbit Flex (ICC = 0.25), and 8.4% Jawbone UP (ICC = 0.50). Cane and walker-user steps underestimated by StepWatch by 1.8% (ICC = 0.98) and 1.3% (ICC = 0.99), respectively, all other monitors underestimated steps by >11.5% (ICCs < 0.05).

ACCEPTED MANUSCRIPT

					Percentage Error	Similar across devices, suggesting the direction of error (underestimation) consistent within each and across all devices.
					Limits of agreement	Wide limits of agreement with larger scatter observed in cane- and walker-user groups for Fitbit One, Omron, Fitbit Flex, and Jawbone UP. Significant variation observed in Omron, Fitbit Flex, and Jawbone UP for cane-users.
Thorup, C. B. et al (2017), Denmark [16]	Shimmer3	Fitbit Zip	waist	steps	Intraclass correlation coefficients (ICC)	Hospitalised patients 24-hour showed relative error of −47.15±24.11 (interclass correlation coefficient (ICC): 0.60), and at home, relative error −27.51±28.78 (ICC: 0.87)
					Percentage Error	Neither 24-hour test had < expected 20% error. During periods of evident walking in 24 hour test, Zip average relative error of <3% at 3.6 km/hour and higher speeds.
Farina, N. et al (2018), UK [17]	Actigraph and NL2000i	Misfit Shine, Fitbit Charge HR	wrist, waist	steps	Intraclass correlation coefficients (ICC)	Two reference devices (ActiGraph and NL2000I) had near perfect agreement (ICC = 0.96, 95% CI: 0.89 to 0.98). All consumer-level activity monitors positively correlated with reference devices (p < .001). Waist-worn Misfit Shine displayed highest agreement amongst devices (ICC = 0.96, 95% 0.91 to 0.99).
					Percentage Error	Misfit Shine (waist) underestimated steps/day vs NL2000i (r_s = 0.90 p<0.001) and overestimated vs ActiGraph device (r_s = 0.96 p <0.001). Misfit Shine (wrist) underestimated steps/day compared to both reference devices (NL2000i r_s = 0.89 and ActiGraph r_s = 0.91 p < 0.001). Fitbit Charge HR (wrist) substantially overestimated step count vs both ActiGraph and NL2000i (NL2000i r_s = 0.83 and ActiGraph r_s = 0.84 p < 0.001).
					Limits of agreement	Misfit Shine (waist) near perfect agreement vs ActiGraph and NL2000i. Compared to ActiGraph and NL2000i, Misfit Shine (Wrist) had good agreement (wide CIs), limits of agreement

		moderately wide. Fitbit Charge HR (wrist) had very wide limits of agreement against ActiGraph and NL2000i.

MVPA = mean vigorous physical activity, 2MWT = two-minute walk test, CI = confidence interval

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 4. Reliability Outcomes

First author, year, location	Consumer-grade activity tracker(s)	Comparison research grade tracker(s)	Sample size (n)	Study Measures	Placement of the consumer-grade activity tracker	Average daily step count recorded by consumer-grade tracker (mean, SD)	Interdevice daily step count variability (consumer vs research-grade tracker)
Paul, S. S. et al (2015), Australia [11]	Fitbit One, Fitbit Zip	ActiGraph	32	steps	waist	NR	Overestimate 716.7 steps/day
Simpson, L. A. et al. (2015), Canada [12]	Fitbit One	ND	42	steps	waist, ankle	N/A	NM
Alharbi, M. et al. (2016), Australia [13]	Fitbit Flex	ActiGraph	48	steps	wrist	NR	Overestimate 1038 steps/day
Boeselt, T. et al. (2016), Germany [14]	Polar A300	Bodymedia Sensewear	20	steps	wrist	NR	Overestimate 183 - 596 (range) steps/day
Floegel, T. A. et al. (2017), US [15]	Fitbit One, Fitbit Flex, Omron HJ-112, Jawbone UP	ActiGraph	99	steps	waist (Fitbit One, Omron HJ-112) and wrist (Fitbit Flex and Jawbone UP)	N/A	NM
Thorup, C. B. et al (2017), Denmark [16]	Fitbit Zip	Shimmer3	24*	steps	waist	7117±5880.6	Underestimate of 2402.7 steps/day
Farina, N. et al (2018), UK [17]	Misfit Shine, Fitbit Charge HR	ActiGraph and NL2000i	25	steps	waist (Misfit Shine) and wrist (Fitbit Charge HR)	NR (ActiGraph 7503.7±3526.3)	Misfit Shine (waist) overestimated 167.6 (MD) steps/ day vs ActiGraph, and underestimated by 633.2 (MD) steps / day vs NL2000i. Misfit Shine (wrist) underestimated 10.1 (MD) steps/day vs ActiGraph and by 899.7 (MD) vs NL2000i.

ACCEPTED MANUSCRIPT

		Fitbit Charge HR (wrist) overestimated 2690.3 steps / day vs ActiGraph and 1721.6 vs NL2000i respectively.

* older participants (not younger healthy controls), MVPA = mean vigorous physical activity, NM = not measured, N/A = not applicable, NR = not reported, MD = mean difference

ACCEPTED MANUSCRIPT