Journal of Science and Medicine in Sport xxx (xxxx)

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Validating the Fitbit Charge 4© wearable activity monitor for use in physical activity interventions

Alexander Waddell [a,*], Stefan Birkett [b], David Broom [a], Gordon McGregor [a,c,d], Amy E. Harwood [b]

[a] Centre for Physical Activity, Sport & Exercise Science, Institute of Health & Wellbeing, Coventry University, UK
[b] Department of Sport and Exercise Sciences, Institute of Sport, Manchester Metropolitan University, UK
[c] Department of Cardiopulmonary Rehabilitation, Centre for Exercise & Health, University Hospitals Coventry & Warwickshire NHS Trust, UK
[d] Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

ARTICLE INFO

Article history:
Received 30 August 2023
Received in revised form 7 December 2023
Accepted 22 January 2024
Available online xxxx

Keywords:
ActiGraph
Accelerometer
Agreement
Reliability
Step count

ABSTRACT

Objectives: Commercially available wearable activity monitors can promote physical activity behaviour. Clinical trials typically quantify physical activity with research grade activity monitors prior to testing interventions utilising commercially available wearable activity monitors aimed at increasing step count. Therefore, it is important to test the agreement of these two types of activity monitors.
Objectives: Observational.
Methods: Thirty adults (20–65 years, n = 19 females) were provided a Fitbit Charge 4©. To determine reliability using an intraclass correlation coefficient, two, one-minute bouts of treadmill walking were performed at a self-selected pace. Subsequently, participants wore both an ActiGraph wGT3X-BT and the Fitbit for seven days. To determine agreement, statistical equivalence and the mean absolute percentage error were calculated and represented graphically with a Bland–Altman plot. Ordinary least products regression was performed to identify fixed or proportional bias.
Results: The Fitbit showed 'good' step count reliability on the treadmill (intraclass correlation coefficient = 0.75, 95 % CI = 0.53–0.87, $p < 0.001$). In free-living however, it overestimated step count when compared to the ActiGraph wGT3X-BT (mean absolute percentage error = 26.02 % ± 14.63). Measurements did not fall within the ± 10 % equivalence region and proportional bias was apparent (slope 95 % CI = 1.09–1.35).
Conclusions: The Fitbit Charge 4© is reliable when measuring step count on a treadmill. However, there is an overestimation of daily steps in free-living environments which may falsely indicate compliance with physical activity recommendations.

Practical implications

- The Fitbit Charge 4© is reliable at measuring step count in a laboratory setting.
- However, it overestimates compared to a research grade activity monitor in a real world setting.
- The Fitbit can therefore be used to promote an increase in step count over time, but should not be used to prescribe specific activity targets.

1. Introduction

Commercially available wearable activity monitors (WAMs) provide real-time data of physical activity behaviour to encourage autonomy and facilitate an active lifestyle.[1] As the wrist being the most user-friendly and accessible location to wear such devices,[2] physical activity wrist watches are frequently used for this purpose. One function of a WAM is measuring step count, which is easily quantifiable.[3] Consequently, WAMs are often used to promote an increase in step count in clinical populations,[4] by providing individuals with an ability to objectively monitor their own activity and take accountability.[5] By helping people to achieve physical activity recommendations, increased step counts facilitated with WAMs may offer health benefits including reduced adiposity and increased insulin sensitivity.[6]

Abbreviations: WAM, wearable activity monitor; ICC, intraclass correlation coefficient; MAPE, mean absolute percentage error; OLP, ordinary least products regression.

* Corresponding author.
E-mail address: alexanderjrwaddell@gmail.com (A. Waddell).
@waddell_1997

https://doi.org/10.1016/j.jsams.2024.01.007

For physical activity interventions, physical activity behaviour is typically quantified by research grade monitors (well established, typically expensive and not available to the general public) before being provided with a WAM to facilitate an increase in step count.[7] It is, therefore, necessary to consider whether WAMs are comparable to research grade activity monitors when quantifying step count[8] — a WAM in agreement with a research grade device would depict similar baseline activity. The Toward Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE) developed recommendations for an optimised validation protocol to ensure consistency of procedures.[9]

The Fitbit Charge 4© is a mid-range wrist worn WAM, capable of a number of physical activity monitoring techniques. As research lags behind product release, evaluation of more recent devices is lacking. Although research trials tend to use older models, as time progresses the prevalence of newer options amongst the general population will only increase. Only one study has evaluated the Fitbit Charge 4© step count function in a laboratory setting in individuals with Huntington's disease[10]; agreement has not been determined in a free-living environment.

Therefore, the aim of this study was to evaluate the reliability of the Fitbit Charge 4© step count function under laboratory conditions, and its agreement in free-living conditions against a commonly used research grade activity monitor, the ActiGraph wGT3X-BT.

2. Methods

The study received ethical approval from Coventry University Research Ethics Committee (P129054). The study sample consisted of 30 adult participants (mean age = 43.3 ± 17.6 years, n = 19 females, mean body mass index = 25.4 ± 4.4 kg/m^2). Recruitment took place using the principal investigator's network and word of mouth within xxx University. The inclusion criteria were adults aged 18 to 65 years, access to a smartphone with Bluetooth capabilities, ability to walk independently and the ability to provide informed consent. Prior to enrolment, potential participants were given an explanation of the investigation, and written informed consent was obtained.

The following two devices were selected for comparison:

1. Commercially available: Fitbit Charge 4© (Fitbit Inc., San Francisco, CA, USA). A wrist-worn WAM with a large organic light-emitting diode screen and a 3-axis accelerometer. Steps are recorded from acceleration data using Fitbit's proprietary algorithm.
2. Research-grade: the ActiGraph wGT3X-BT (ActiGraph LLC, Pensacola, FL, USA) activity monitor that captures and records raw data at a sample frequency between 30 and 100 Hz using a 3-axis accelerometer. The raw data is then post-processed into epochs using the ActiLife software. Step counts are generated on a per epoch basis using data collected on the vertical axis.

The INTERLIVE network provides recommendations for assessing the validity and reliability of devices.[9] In the context of a physical activity clinical trial, we deemed it only necessary to demonstrate that the Fitbit Charge 4© is reliable at recording steps, and that it is valid when compared to an ActiGraph wGT3X-BT in free living conditions. Therefore, only the relevant aspects of the recommendations were followed, as described below. After providing informed consent, sociodemographic and anthropometric information was collected including age, sex, height, and body mass. The participants were then given a Fitbit Charge 4© to wear on the non-dominant wrist (manufacturer instructions), and the device was set up using a study specific email address.

To assess intra-device reliability of the Fitbit Charge 4©, participants were asked to complete two bouts of one-minute, steady state walking at a self-selected pace on a treadmill.[9] The number of steps recorded during each bout was recorded and compared. To calculate the number of steps recorded during each minute of walking, participants were asked to stand still at the start and end of each walking bout, to allow the researcher to note the number of steps that had been recorded during each walk. The treadmill was paused and resumed between bouts to ensure the same speed was utilised.

Validity was determined by comparing the step recordings between the Fitbit Charge 4© and an ActiGraph wGT3X-BT over one week in a free-living environment. Following the end of the treadmill assessment, participants were set up with the ActiGraph wGT3X-BT activity monitor to wear around the waist on the non-dominant side (manufacturer instructions). Participants were asked to wear both devices continuously, only removing them when washing or sleeping, and to synchronise the Fitbit with the Fitbit app daily. A physical wear log was completed to record instances where devices were removed, and both devices were returned after one week. Participants verbally confirmed that they had undertaken a typical/'normal' week and that they had not experienced any injuries that would have influenced gait.

Fitbit and ActiGraph data were downloaded and total step count for the week and average daily steps were determined. For ActiGraph data, the sampling frequency was set at 30 Hz, without the low frequency extension applied (firmware version 1.9.2). Data were downloaded using the ActiLife software (version 6.13.4) with 60 second epochs, calculating daily steps automatically. Fitbit accelerometer raw data can be obtained using an application programming interface (API), however this was not within the scope of the present study. Step count data was therefore obtained from each participant's user dashboard. Fitbit data by default is collected at a sampling frequency of 100 Hz. Valid wear time for both the Fitbit and ActiGraph wGT3X-BT was defined as a minimum of three days of 10 h of wear time, including at least one weekend day.[11] ActiGraph non-wear time was defined by a 90-minute time window of consecutive no counts and no interruptions.[11,12] Participants were instructed to wear both devices simultaneously, recording the removal of both on the wear log. Discrepancies in non-wear time between the ActiGraph wGT3X-BT and Fitbit were determined on return of the wear log.

All quantitative statistics are presented as means and standard deviations unless otherwise stated. Descriptive statistics for sociodemographic information were calculated. All data was assessed for normality both visually via histograms and with the Shapiro Wilk test. The level of agreement between the Fitbit and ActiGraph was assessed with Bland–Altman limits of agreement, and with standardised tests of statistical equivalence.[13] The null hypothesis of a difference between the Fitbit and ActiGraph step counts was tested with a pre-determined equivalence region of ± 10 % of the mean of the ActiGraph steps (based on the conservative recommendations of an acceptable MAPE of 10–15 % in free living conditions[9]). The test was performed to determine if the 90 % confidence intervals of the difference between the two measures fell within the equivalence zone with 95 % precision. An ordinary least-product (OLP) regression was calculated to determine the degree of fixed/proportional bias.[14,15] Differences between the two devices were determined using the mean absolute percentage error (MAPE). An intraclass correlation coefficient (ICC) was performed for the two treadmill walking bouts to quantify the intra-device reliability of the Fitbit. The ICC estimate and 95 % confidence intervals were calculated based on a single rating, absolute agreement, two-way mixed effects model. An ICC of >0.9, 0.75–0.90, 0.50–0.75, and <0.50 was classified as excellent, good, moderate, and poor, respectively.[16] All statistical analyses were performed using the latest version of R (4.1.3).[17] The alpha level was set at 0.05. The study data and analysis code are available in the Supplementary material.

3. Results

All 30 participants completed study procedures. Participant characteristics are shown in the Supplementary material.

Each monitored day was deemed valid for analysis, as the average wear time for all participants was 824.45 ± 142.44 min per day, exceeding the required 10 h. No discrepancies were observed on the wear log between the two devices, meaning wear time can be interpreted as similar.

Laboratory evaluation of the Fitbit reliability – through repeated treadmill walking – revealed a 'good' ICC of 0.75 (95 % CI = 0.53–0.87, $p < 0.001$).

Average step counts after one week of free-living for both devices are shown in Table 1. The MAPE between the Fitbit and ActiGraph for total weekly steps was 26.02 % (95 % CI = 20.78–31.26). The absolute percentage error ranged from 0.87 % to 58.32 %.

OLP regression revealed the Fitbit showed proportional bias, whereby it was consistently measuring a higher number of steps compared to the ActiGraph, at a level proportional to the number of steps measured in total (i.e. the more steps that were recorded, the greater the degree of error). This was demonstrated by the 95 % confidence intervals of the slope not including 1 (see Supplementary material).[14]

The Bland–Altman plot for daily average step count is shown in Fig. 1. Given the OLP regression demonstrated proportional bias, the Bland–Altman plot was constructed using an ordinary least squares regression line of best fit and hyperbolic 95 % confidence limits.[18] Wide limits of agreement indicate a high individual predictive error, and the plots show a trend of Fitbit's overestimating steps in all participants compared to the ActiGraph.

The equivalence zones are shown in Table 1 and Fig. 2. The equivalence region was predetermined as ±10 % of the ActiGraph mean weekly steps based on the conservative recommendation from INTERLIVE (10–15 % in free living environments).[9] The 90 % confidence intervals of the difference between the ActiGraph and Fitbit do not fall within the equivalence region, meaning the null hypothesis is unable to be rejected, and thus the Fitbit cannot be considered equivalent to the ActiGraph for recording steps ($p = 0.999$).

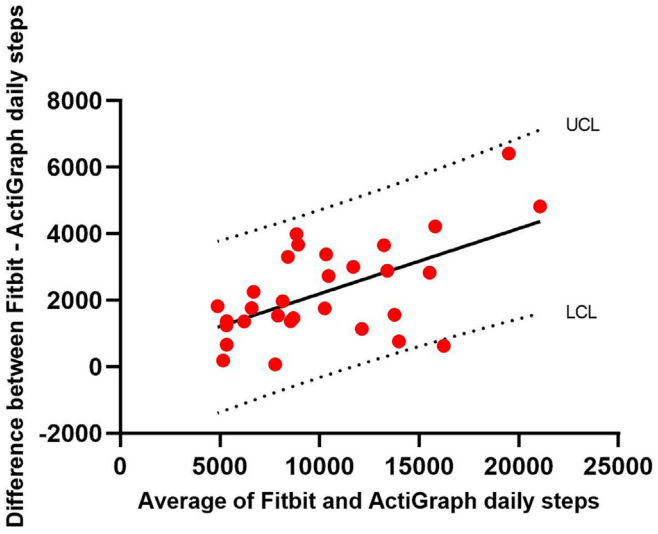

Fig. 1. Bland–Altman plot of differences against averages for Fitbit & ActiGraph daily steps over one week. Solid line represents ordinary least squares regression line of best fit. UCL; upper 95 % confidence limit, LCL; lower 95 % confidence limit.

4. Discussion

The purpose of this study was to determine the reliability of the Fitbit Charge 4©'s step count function in a laboratory setting, and evaluate agreement in a free-living environment compared to a research grade activity monitor to assess its suitability for use in a clinical research trial.[19] In agreement with previous research (ICC of 0.81),[10] we report 'good' reliability for the Fitbit Charge 4© step count function (ICC of 0.75) on a treadmill at self-selected walking speeds. However, the Fitbit consistently overestimated step count compared to an ActiGraph wGT3X-BT.

In contrast to our findings of overestimation, Curran et al[20] demonstrated that the Fitbit Charge 2© underestimated steps compared to direct observation in a laboratory setting. One explanation is that movement is more irregular/intermittent in free-living compared to in a laboratory, which has shown to increase the error rates of WAMs.[21,22] We have also shown that the Fitbit Charge 4© displays proportional bias, indicating, that error increases the more steps are recorded. Research limited to assessing validity solely in laboratory settings via a treadmill test will record far fewer total step counts than those over an extended period of time in free-living conditions, providing another potential explanation for why underestimation is seen in a laboratory whereas overestimation is seen in free-living environments. This is consistent with a review of the literature, which found a range of Fitbits to underestimate in laboratory conditions and overestimate in free-living conditions.[23] We have demonstrated that this is also the case with a more recent model, indicating a consistent pattern.

In contrast, Burton et al[24] had participants wear devices for 14 days of free-living, but showed underestimation of step-count with a Fitbit. However, a different device (GENEactiv) was used as the comparator, thus preventing direct comparison to our study using the ActiGraph. A potential explanation for why overestimation was not observed in this case is that walking speed is a key determinant of step-count in wrist-worn activity monitors.[23,25] The associated algorithms for determining a step may lack accuracy at lower speeds due to the limited acceleration data received. As participants in the study by Burton et al[24] were older, walking speed may have differed, thus lowering the number of steps recorded at the wrist. In our study, we included participants ranging from 20 to 65 years of age, where walking speeds will naturally vary considerably. Despite this, the Fitbit consistently overestimated steps in every participant, as shown in the Bland–Altman plots.

An analysis of wear location for the ActiGraph GT3X and Fitbits showed that compared to the waist, wrist worn devices in free-living resulted in significantly more steps being recorded.[23,26] A potential explanation may be that during sedentary activities wrist kinematics represent a large proportion of total body movement, leading to incorrect recognition of steps by the WAM sensors.[27] Additionally, devices worn at the hip may be more accurate as they are closer to the body's centre of mass and so better represent whole body acceleration than at the wrist.[26] Further, across a range of speeds, the waist appears to have the lowest overall error compared to actual visually recorded step count.[28] An ActiGraph around the waist is therefore a more appropriate criterion in free-living conditions, where one may be walking, doing general activities, or undertaking a range of exercise modes. We required the wGTX3-BT to be worn at the waist rather than the wrist, as this is commonly adopted methodology when quantifying habitual physical activity behaviour in clinical trials. There is therefore a discrepancy between commonly adopted methodology for research trials which have waist-worn devices due to the increased accuracy, and wrist worn devices used by the general public due to the increased convenience.

There is no widely accepted error margin cut-off for confirming device validity.[9] The INTERLIVE network recommends that device validity be assessed within the context of the intended use. For example, if a device is to be used as an objective outcome of physical activity in a clinical trial, or under controlled laboratory conditions, the acceptable MAPE should be low (<5 %). If the device is intended for use by the general

Table 1
Weekly step counts for ActiGraph wGT3X-BT and Fitbit Charge 4© and equivalence testing statistics.

Statistic	Value
ActiGraph weekly steps, mean (SD)	66,908 (29398)
Fitbit weekly steps, mean (SD)	83,455 (35735)
Difference in weekly steps, mean (SD)	16,475 (10800)
10 % equivalence region	−6698, +6698 or 60,282–73,768
90 % confidence intervals of the difference	−13,125.28, +19,825.72 or 53,854.72–86,805.72

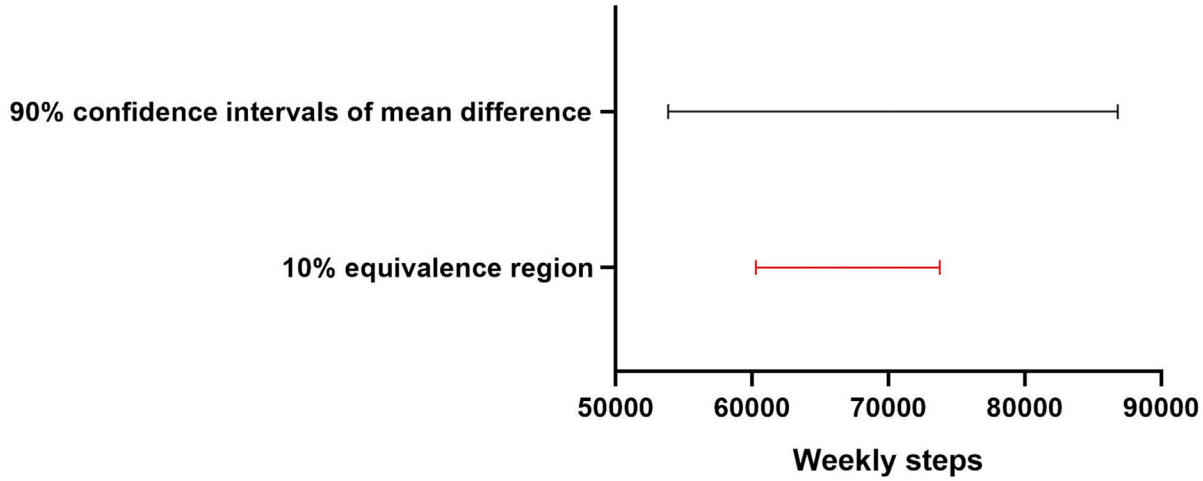

Fig. 2. Equivalence testing shows the Fitbit Charge 4© is not equivalent to the ActiGraph wGT3X-BT for measuring step count as the 90 % confidence intervals of the difference between the two measures do not fall within the ±10 % equivalence region (shown in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

population, an acceptable MAPE may be 10–15 %.[9] Given our findings show a MAPE greater than even the upper limit of this suggested error range, the Fitbit Charge 4© could be seen to have poor agreement with the ActiGraph wGT3X-BT.

An issue with using a WAM that overestimates steps is that clinical populations and practitioners may be under false pretenses that step goals associated with better health outcomes are being achieved.[29] This may prevent people from surpassing these thresholds due to feeling satisfied once achieving their goal. If the WAM is overestimating, therefore, people may not gain any clinically meaningful benefit. However, whilst a WAM may overestimate, it can still be reliable (i.e. consistently over or underestimating step count). If this is the case, as long as step count is increasing consistently, the critical factor of promoting autonomy and behaviour change may still be achieved. The same can be said for proportional bias: If the main point of interest is increasing step count, as long as it consistently increases, the objective will still have been achieved.[14] Therefore, in a clinical scenario when prescribing steps, increases over time should be the focus, rather than achievement of a specific target. However, if a study's aims are to detect between group differences in step counts, devices should have a high degree of agreement.

5. Strengths and limitations

The strengths of this study are that step count assessments were performed in both controlled laboratory and free-living environments. This involved both walking on a flat treadmill as well as over a range of terrains that would be experienced whilst ambulating in the free-living. The participants were of a wide range of ages and heights, which naturally resulted in a spread of self-selected walking speeds and step length both in the laboratory and free-living environments. This helped to provide a sample representative of the general population (age range = 18–65 years, BMI = 18.5–37.9 kg/m^2).

Comparing our data with existing literature on device agreement can be problematic due to the use of different criterion measures, older or different WAM models and varying study design. However, as the overall aims of the study surround WAM applications in research, the trends of wear location/walking speed/laboratory vs free living etc. are still warranted. To allow our data to be comparable against future studies, we followed the standardised INTERLIVE recommendations.[9] All participants were healthy adults, so the results cannot be generalised to clinical populations, which may be in greater need of physical activity behaviour monitoring. Only individuals who could walk independently were included, meaning we were unable to discern error rates in those with walking aids or gait disturbances, although we expect the MAPE would be even higher. Laboratory evaluation was also only completed for two periods of 1 min — walking for longer may have returned different results. Additionally, self-selected walking speeds were not recorded, thus limiting analysis of the effects of speed. As the aim was to quantify reliability, and assess agreement in free-living only, during laboratory testing, no gold standard criterion such as video recording was performed to evaluate agreement in this context. Finally, as an API was not utilised to analyse Fitbit data, more detailed comparisons (i.e. on a minute-by-minute level) were not possible, and sampling rate could not be altered, and so differed between devices.

6. Conclusion

As a result of our findings regarding the reliability of the Fitbit Charge 4© step count function supported by previous work,[10] researchers may want to consider this device when conducting a physical activity intervention focused on increasing step count. They are widely available, relatively cheap and are wrist worn, making them easily accessible and non-invasive.[30] Real-time feedback has the added potential for promoting physical activity behaviour change. However, if the aim is to utilise a research grade device such as the wGT3X-BT to quantify habitual step count, before then using a Fitbit as part of an intervention, this overestimation should be considered. A potential solution may be to first determine habitual step count with a research grade device (in which participants are blind to the data), and then to prescribe physical activity targets taking into account the MAPE of the Fitbit Charge 4©. For example, if participants' ActiGraph showed they were averaging 5000 steps a day at baseline, to account for the 26.02 % MAPE of the Fitbit, initial goals could be prescribed on the Fitbit at 6300 steps. Another alternative may be to use research grade accelerometers for both purposes. The ActiGraph GT9X is wrist-worn and has a display, eliminating the need to use different devices with separate wear locations and algorithms. However, allowing participants to visualise their behaviour prior to an intervention may not provide a true representation of habitual physical activity behaviour. Future research should evaluate other physical activity metrics of the Fitbit Charge 4© (i.e. sleep, sedentary behaviour etc.).

Given the global reach of Fitbit, the push from the World Health Organization for people to be more physically active and the fact wearable technology continues to top the ACSM's worldwide fitness trends, the general public should also be aware of these findings. Setting specific step count goals based on suggestions from governing bodies will likely not result in any benefit if in reality they are consistently fallen short of.

Funding information

Funding for this project was included as part of a fully funded PhD for the principal investigator (AW) with the Centre for Physical Activity, Sport and Exercise Sciences, Research Institute of Health and Wellbeing, Coventry University.

Confirmation of ethical compliance

This study was given ethical approval by the Coventry University Research Ethics Committee (P129054). This study was conducted in accordance with the principles of the Declaration of Helsinki.

CRediT authorship contribution statement

Alexander Waddell: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Stefan Birkett:** Conceptualization, Writing – review & editing. **David Broom:** Conceptualization, Writing – review & editing. **Gordon McGregor:** Conceptualization, Writing – review & editing. **Amy E. Harwood:** Conceptualization, Writing – review & editing.

Declaration of interest statement

The authors declare no conflicts of interest.

Acknowledgements

None to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsams.2024.01.007.

References

1. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 2015;313:459-460.
2. Alley S, Schoeppe S, Guertler D et al. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. *BMJ Open* 2016;6:e011243.
3. Bassett DR, Toth LP, LaMunion SR et al. Step counting: a review of measurement considerations and health-related applications. *Sports Med* 2017;47:1303-1315.
4. Ferguson T, Olds T, Curtis R et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. *Lancet Digit Health* 2022;4:e615-e626.
5. Gualtieri L, Rosenbluth S, Phillips J. Can a free wearabl. e activity tracker change behavior? The impact of trackers on adults in a physician-led wellness group. *JMIR Res Protoc* 2016;5(4):e237. https://www.researchprotocols.org/2016/4/e237. [2016; 5: e6534].
6. Dwyer T, Ponsonby AL, Ukoumunne OC et al. Association of change in daily step count over five years with insulin sensitivity and adiposity: population based cohort study. *BMJ* 2011;342:216.
7. Kim Y, Lumpkin A, Lochbaum M et al. Promoting physical activity using a wearable activity tracker in college students: a cluster randomized controlled trial. 2018;36: 1889-1896. doi:10.1080/02640414.2018.1423886.
8. Brickwood K-J, Watson G, O'brien J et al. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019;7:e11819.
9. Johnston W, Judice PB, Molina García P et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med* 2021;55:780-793.
10. Mann E, Drew CJ, Busse M et al. F62 Validation of digital assessment of physical activity in Huntington's disease: comparing fitbit charge 4 step count with research-grade accelerometers. *J Neurol Neurosurg Psychiatry* 2022;93:A58-A59.
11. Migueles JH, Cadenas-Sanchez C, Ekelund U et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Med* 2017;47:1821-1845.
12. Peeters G, van Gellecum Y, Ryde G et al. Is the pain of activity log-books worth the gain in precision when distinguishing wear and non-wear time for tri-axial accelerometers? *J Sci Med Sport* 2013;16:515-519.
13. Dixon PM, Saint-Maurice PF, Kim Y et al. A primer on the use of equivalence testing for evaluating measurement agreement. *Med Sci Sports Exerc* 2018;50:837.
14. Ludbrook J. Comparing methods of measurements. *Clin Exp Pharmacol Physiol* 1997;24:193-203.
15. Nascimento-Ferreira M, de Moraes A, Vinicius Nascimento-Ferreira M et al. *Ordinary Least Products Regression: A Robust Statistical Tool for Assessing Agreement Between Measures Attended by High Variability*, September 10, 2020. doi:10.21203/RS.3.RS-71462/V1. [Epub ahead of print].
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-163.
17. R Core Team. *R: A Language and Environment for Statistical Computing*, 2022.
18. Ludbrook J. Confidence in Altman–Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol* 2010;37:143-149.
19. Waddell A, Denton F, Powell R et al. Community WALKing and home-baSed circuiT tRaining in peOple liviNG with intermittent claudication (WALK-STRONG): protocol for a randomised controlled feasibility trial. *J Vasc Soc GB Irel* 2022;1. doi:10.54522/JVSGBI.2022.24. [Epub ahead of print].
20. Curran M, Tierney A, Collins L et al. *Accuracy of the ActivPAL and Fitbit Charge 2 in Measuring Step Count in Cystic Fibrosis*, 2021. doi:10.1080/09593985.2021.1962463. [Epub ahead of print].
21. Hickey A, John D, Sasaki JE et al. Validity of activity monitor step detection is related to movement patterns. *J Phys Act Health* 2016;13:145-153.
22. Toth LP, Park S, Pittman WL et al. Effects of brief intermittent walking bouts on step count accuracy of wearable devices. *J Meas Phys Behav* 2019;2:13-21.
23. Feehan LM, Geldman J, Sayre EC et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth Uhealth* 2018;6(8): e10527. https://mhealth.jmir.org/2018/8/e10527.
24. Burton E, Hill KD, Lautenschlager NT et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. *BMC Geriatr* 2018;18:1-12.
25. Fokkema T, Kooiman TJM, Krijnen WP et al. Reliability and validity of ten consumer activity trackers depend on walking speed. *Med Sci Sports Exerc* 2017;49:793-800.
26. Mandigout S, Lacroix J, Perrochon A et al. Comparison of step count assessed using wrist- and hip-worn Actigraph GT3X in free-living conditions in young and older adults. *Front Med (Lausanne)* 2019;6:252.
27. Aziz O, Robinovitch SN, Park EJ. Identifying the number and location of body worn sensors to accurately classify walking, transferring and sedentary activities. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS* 2016:5003-5006. [2016-October].
28. Karaca A, Demirci N, Yılmaz V et al. Validation of the ActiGraph wGT3X-BT accelerometer for step counts at five different body locations in laboratory settings. 2021;26:63-72. doi:10.1080/1091367X.2021.1948414.
29. Jayedi A, Gohari A, Shab-Bidar S. Daily step count and all-cause mortality: a dose–response meta-analysis of prospective cohort studies. *Sports Med* 2022;52:89-99.
30. Shelley J, Fairclough SJ, Knowles ZR et al. A formative study exploring perceptions of physical activity and physical activity monitoring among children and young people with cystic fibrosis and health care professionals 11 Medical and Health Sciences 1117 Public Health and Health Services. *BMC Pediatr* 2018;18:1-16.