Published in final edited form as:

Physiol Meas. ; 41(1): 015001. doi:10.1088/1361-6579/ab635f.

Validity of Smartphones and Activity Trackers to Measure Steps in a Free-living Setting over Three Consecutive Days.

Christoph Höchsmann^{1,2}, Raphael Knaier¹, Denis Infanger¹, Arno Schmidt-Trucksäss¹

¹Department of Sport, Exercise and Health, University of Basel, Basel, Switzerland ²Pennington Biomedical Research Center, Baton Rouge, LA, USA

Abstract

Objective: This study assessed the validity of a consumer activity wristband, a smartphone, and a research-grade accelerometer to measure steps in a free-living setting.

Approach: Thirty healthy adults were equipped with two Garmin Vivofit (non-dominant wrist), one iPhone SE (right pants pocket), three ActiGraph wGT3X+ (two on the hip, one on the non-dominant wrist), and one StepWatch (right ankle) and instructed to wear the devices continuously during a 3-day monitoring period. All activities of daily living were recorded in 15-minute intervals in a diary. The StepWatch served as the criterion method and validity was evaluated by comparing each device with the criterion measure using mean absolute percentage errors (MAPE).

Main results: The MAPE for the total step count during the 3-day monitoring period was high with a general underestimation of steps by all devices of >20% compared to the criterion measure. The wrist-worn ActiGraph markedly overestimated steps during predominantly low active (public transport or driving, and standing) or even inactive (sitting and lying) activities of daily living.

Significance: The overall high MAPE of the devices compared to the StepWatch during step-based activities, likely caused by inaccuracies during short and intermittent bouts of activity, may limit their validity in a free-living setting.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere. As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period. After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>. Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record. View the [article online](#) for updates and enhancements.

Corresponding author: Prof. Arno Schmidt-Trucksäss, MD, Division of Sports and Exercise Medicine, Department of Sport, Exercise and Health, University of Basel, Birsstrasse 320 B, 4052 Basel, Switzerland, Phone: + 41 61 207 47 40, Fax: + 41 61 207 47 42, arno.schmidt-trucksass@unibas.ch.

Authors' contributions: CH, RK, and AST conceived the study and designed the methods. CH and RK coordinated the study and collected data. DI, CH, and RK performed the statistical analysis. CH and RK drafted the manuscript, DI drafted the figures. All authors interpreted results, commented on drafts and approved the final version of this manuscript.

Ethics approval and consent to participate: This study was approved by the ethics committee "Ethikkommission Nordwest- und Zentralschweiz", Switzerland (EKNZ 2017-01819) and complied with the declaration of Helsinki. Written informed consent was obtained from participants before participation in the study.

Competing interests: The authors declare that they have no competing interests.

Keywords

accelerometer; physical activity measurement; step count; ambulatory activity

INTRODUCTION

Measurement of physical activity (PA) through accelerometers or pedometers is of high relevance in many research settings, and high wear-time compliance of the devices is indispensable for an accurate assessment¹. Hip-worn research-grade accelerometers are often not worn reliably enough to ensure sufficient wear-time compliance and consequently do not allow an accurate PA assessment². In recent years, wrist-worn activity trackers have gained popularity, and their wear-time compliance seems to be higher compared to hip-worn accelerometers because of their greater wearing comfort and because participants are less likely to remove them at night or during the day or forget to wear them at all¹⁻³. In addition to wrist-worn activity trackers, smartphones with their built-in accelerometers have increasingly gained attention with regard to PA tracking. Provided that sufficient validity in free-living conditions is confirmed, the fact that approximately two billion people worldwide use smartphones daily⁴ makes them potentially excellent tools for PA assessment in large populations, as additional devices for PA monitoring would become unnecessary if a sufficient wear-time is ensured.

In a first validation study⁵, we were able to confirm the validity of a popular activity wristband (Garmin Vivofit 2) and two popular smartphones (Apple iPhone SE and Samsung Galaxy S6) to detect steps during different walking speeds on a treadmill and during an outdoor walking course at a self-chosen walking speed. The results of that study support previous findings⁶⁻¹⁰ and extend them^{8,9} by showing that the validity of step detection is independent of the phone's placement on or near the body. However, the devices tested in previous studies were either not validated in a free-living setting^{5,7,9} or if so, not during several days^{6,10,11}, or against devices that are known to have poor validity during slow walking speeds^{12,13}. Because slow walking and incidental, non-continuous ambulation (activities of daily living) make up large parts of the overall daily PA, validation of accelerometers in a free-living setting is crucial. In addition, because one day often does not represent real-life conditions, it is highly advisable to validate devices over several days. In general, accelerometers are known to underestimate steps during short activity bouts in which only a few steps are taken as well as during slow walking speeds¹⁴. Further, particularly wrist-worn devices tend to record false steps during free-living conditions due to arm movements during activities without ambulation, leading to an overestimation of steps^{15,16}. Both factors can significantly influence the accuracy of the PA assessment, underlining the importance of validating the devices during those everyday conditions.

Finally, in addition to validity, reliability of accelerometers is imperative if they are to be used in research or clinical settings because the degree of inter-instrument reliability will inevitably influence the device's ability to detect changes in PA behavior^{17,18}. While the reliability of research-grade accelerometers has been shown before under laboratory- and

free-living conditions^{19,20}, inter-instrument reliability data for consumer wristbands under free-living conditions are scarce in general and do not exist for the Garmin Vivofit 2.

Therefore, the aim of this project was to compare accelerometer devices of different tiers with regards to recorded step counts in a free-living setting over three consecutive days. Specifically, we compared (1) the StepWatch, a research-grade, ankle-worn device, (2) the research-grade ActiGraph wGT3X+ accelerometer (worn on the hip and wrist), (3) the Garmin Vivofit 2, a consumer activity wristband, and (4) two types of smartphones (iPhone SE and Samsung Galaxy S6 Edge). To quantify the measurement error of the different devices, we compared the step counts as recorded by the ActiGraph wGT3X+, Garmin Vivofit 2, and smartphones with those of the StepWatch, which we used as reference method due to its repeatedly confirmed validity during various walking conditions and for different populations^{11,21–23}. The aim was further to detect the recording of “false” steps of the devices (i.e., steps recorded during non-ambulatory activities) and to assess the accuracy of the devices for different activities of daily living. Finally, we aimed to assess and compare the reliability of the Garmin Vivofit 2 and ActiGraph wGT3X+ to evaluate the devices’ stability in measuring steps in a free-living setting over three days.

METHODS

Participants

Thirty healthy adults were recruited from the University of Basel, Switzerland, and the surrounding community through posting/flyers within the University of Basel and by word of mouth. Participants with orthopedic or other physical limitations that may impair daily PA were excluded from this study. Eligible participants were provided with detailed information about the purpose and procedures of this study, completed a health history questionnaire, and gave written informed consent before participation.

Procedures

All experimental procedures of this study were reviewed and approved by the local ethics committee (EKNZ 2017–01819) and performed in accordance with the Declaration of Helsinki. Data were collected between December 2017 and July 2018.

Participants were provided with the following accelerometer devices and instructed to wear them in standardized positions on the body during the next three days while awake. For all devices, sex, age, height, and weight were entered into the respective software.

Garmin Vivofit 2—Two Garmin Vivofit 2 (software version 3.5) devices (Garmin International Inc., Olathe, KS, USA) were worn on the non-dominant wrist as recommended by the manufacturer. One device was placed just above the wrist bone and the other one slightly more proximal above the ActiGraph (see below).

ActiGraph wGT3X+—The ActiGraph wGT3X+ (ActiGraph LLC, Pensacola, FL, USA) was placed on the non-dominant wrist in between the two Garmin Vivofit devices as well as on the opposite hip (two devices directly next to each other). Both placements are recommended by the manufacturer. The sampling rate was set to 60 Hz and ActiLife v6.13.3

was used to assess step data post hoc with the epoch length set to 60 seconds. Low-frequency extension (LFE) was disabled, as LFE processing of raw data has been shown to lead to an overestimation of actual steps due to a greater amount of movement artifacts being counted as steps²⁴, particularly when the device is placed on the wrist¹¹.

Apple iPhone SE—The iPhone SE (Apple Inc., Cupertino, CA, USA) includes a three-axis gyroscope and accelerometer as well as the M9 motion coprocessor and it was placed in the front pocket of the pants on the same side as the hip-worn ActiGraph devices. Steps were recorded with the built-in Apple Health app (iOS version 11.2).

Samsung Galaxy S6 Edge—The Samsung Galaxy S6 Edge (Samsung Electronics Co., Ltd., Suwon, South Korea) is equipped with a three-axis gyroscope and accelerometer and it was placed in the front pocket of the pants on the opposite side of the iPhone. Steps were recorded with the built-in Samsung S Health app (Android version 8.1). This phone was used in a subsample (n=9) only.

StepWatch 3—The StepWatch 3 (Modus Health Inc., Washington, DC, USA) was placed on the lateral side of the right ankle, directly above the malleolus and affixed with elastic straps, as suggested by the manufacturer. Before attaching the StepWatch to the ankle, each device was initialized using the “Easy Start” menu and default sensitivity and cadence setting within the Modus software (version 3.4). The default setting has been shown to be highly accurate during walking and light, intermittent lifestyle activities in healthy, able-bodied adults²⁵. We used the StepWatch as the criterion measure in this study. The StepWatch has shown excellent validity in laboratory²⁶ and free-living¹¹ conditions and has been used as criterion measured in previous validation studies²².

Wear time validation—Participants recorded the wear time of all devices as well as the times awake for each day in a diary. Participants were instructed to place the devices on their body directly after getting up in the morning and to wear them continuously throughout the day until they went to bed. If the devices had to be taken off throughout the day, participants were further instructed to always put on and take off all devices at the same time.

Recording of activities of daily living—Participants recorded all activities of daily living (step-based activities, cycling, public transport (and driving), standing, sitting, and lying) in 15-minute intervals in a diary throughout the study period. The step counts as recorded by the different devices during these intervals were then summed per activity and averaged per day. We chose the timeframe of three days as it seemed reasonable to expect that all typical activities of daily living would be performed by our participants during that timeframe if they were, in fact, typical activities. A longer timeframe would certainly have provided even more data but would also likely have affected participants' compliance in accurately recording all activities in the diary.

Statistical analyses

Baseline characteristics of the participants were summarized using the median and interquartile range (IQR). Reliability of the devices was assessed by calculating the

Intraclass Correlation Coefficient (ICC; two-way random, absolute agreement) and corresponding 95% confidence intervals (CI). In addition, we calculated the mean absolute percentage error (MAPE) to compare each device with the criterion measure (StepWatch). Bland-Altman analyses were used to evaluate the level of agreement between the StepWatch and all other devices with regard to steps per day. It was not possible to include the smartphones in the detailed analysis of step counts during different activities of daily living, as we were not able to access the recorded data in intervals <60 minutes. We used R 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) for our analyses and graphics.

Sample Size—Due to insufficient data regarding the validity of our devices to detect steps in a free-living setting, we calculated the sample size based on the expected ICC²⁷. A systematic review of comparable validation studies has shown that the ICC is usually well above 0.7²⁸. With a significance level of 0.05, the sample size needed to attain a conservatively calculated ICC of 0.5 at a targeted power of 80% was determined as 28 participants. Accounting for a possible dropout of 10%, we aimed to include 30 participants in our study.

RESULTS

Thirty participants were included in this validation study (Table 1) and completed the 3-day monitoring period. One participant was excluded due to abnormally high levels of PA (27,900 steps per day) and large amounts of vigorous-intensity PA throughout the 3-day monitoring period, which may have conflicted with the selected settings for the StepWatch and consequently may have impacted the accuracy of the device in that participant²⁵. Therefore, 29 participants were included in the main analyses. Wear time of the devices was high with a mean of 800 (standard deviation (SD) 128) minutes per day equating to 86.7% (SD 10.8 %) of the total time awake with identical wear times of all devices in the same participant. The StepWatch recorded a mean of 9,723 (SD 3,131) steps per day (minimum 4,772; maximum 17,065), classifying our sample as active²⁹ on average.

Step detection of the devices compared to the StepWatch

The ICC was above 0.5 for all devices and therefore higher than assumed in our sample size calculation.

Figure 1 shows the mean difference in measured steps per day of all devices compared to the criterion measure. Throughout the 3-day monitoring period, the MAPE for the different devices compared to the criterion measure was 27% for the iPhone, 27% for the Samsung Galaxy, 22% (distal wrist) and 23% (proximal wrist) for the Garmin Vivofit, 28% (ActiGraph 1) and 29% (ActiGraph 2) for the hip-worn ActiGraph devices, and 14% for the wrist-worn ActiGraph.

The Bland-Altman plots (Figure 2) show the systematic underestimation of steps as measured by the iPhone SE, Garmin Vivofit (both placements) and hip-worn ActiGraph wGT3X+ compared to the StepWatch. The wrist-worn ActiGraph wGT3X+ deviated unsystematically from the StepWatch, tending to overestimate steps at lower daily step values and to underestimate at higher daily step values.

Accuracy of the devices during different activities of daily living

Figure 3 illustrates the average number of steps as recorded by the different devices during typical diary-assessed everyday activities. During step-based activities, a mean systematic underestimation of steps by the Garmin Vivofit and ActiGraph devices becomes visible. On average, the Garmin Vivofit underestimated daily steps during step-based activities by 1,604 (SD 1,003) steps per day or 19% when worn on the distal wrist and by 1,749 (SD 1,089) steps per day or 21% when worn on the proximal wrist. The hip-worn ActiGraph showed an average underestimation of 2,339 (SD 1,337) (ActiGraph 1) and 2,395 (SD 1,355) (ActiGraph 2) steps per day during these activities (equating to 28% and 29%), and the wrist-worn ActiGraph underestimated steps by 1,069 (SD 925) steps per day (13%).

During cycling, the StepWatch (criterion measure) recorded an average of 1,047 (SD 1,064) steps per day. All other devices recorded step counts that were on average 55% lower than those recorded by the criterion measure.

During everyday activities that are predominantly low active (public transport or driving, and standing) or even inactive (sitting and lying), the wrist-worn ActiGraph markedly overestimated steps per day compared to the criterion measure with a deviation of 57% during standing and as much as 215% during sitting. All other devices are comparable to the criterion measure.

Reliability of the devices

Figure 4 illustrates the absolute difference both Garmin Vivofit (proximal and distal wrist) and both hip-worn ActiGraph devices. The Garmin Vivofit worn on the distal wrist recorded 3% more steps than the device that was worn on the proximal wrist with a mean absolute difference of 210 (SD 204) steps per day. The absolute difference between both hip-worn ActiGraph devices was 130 (SD 102) steps per day, equal to a 2% deviation.

DISCUSSION

This validation study assessed the accuracy of two types of smartphones, one consumer activity wristband, and a research-grade accelerometer at two different placements (hip and wrist) to measure steps during typical everyday activities in a free-living setting over three days compared to the ankle-worn StepWatch. Overall, all devices markedly underestimated daily steps compared to the criterion measure, leading to MAPE values of >20%. In people who primarily walk and perform light, intermittent lifestyle activities such as our study sample, the StepWatch with default settings has been shown to have good validity under free-living conditions with a slight overestimation (MAPE 4.1%) of actual, hand-counted steps¹¹.

The MAPE for the hip-worn ActiGraph (28%–29%) in our study was comparable to that of the same device in another free-living validation study, which reported an underestimation of steps of around 25% compared to the StepWatch across one full day¹¹. The Garmin Vivofit likewise showed MAPE values that are comparable to other popular consumer-grade devices with an average underestimation of around 20% of the criterion measure¹¹. As the accuracy of the Garmin Vivofit during various continuous (even very slow) walking speeds has been

confirmed before⁵, it is likely that the underestimation of steps by the device mainly occurred during intermittent walking bouts. This assumption is supported by findings that show that the Garmin Vivofit 2, much like most consumer activity wristbands, undercounts steps during brief, intermittent walking bouts of eight steps or less, as the device uses an algorithm that detects continuous, rhythmic ambulation and filters out bouts of ambulation that last only a few seconds³⁰. In this study, it is likely that typical household activities that are characterized by intermittent ambulation and that were correctly recorded, as step-based everyday activities in the diary were primarily responsible for the underestimation of steps.

All devices recorded steps during activities that were marked as low active (public transport and standing) and even inactive (sitting and lying) activities in the diaries. For the Garmin Vivofit and the hip-worn ActiGraph, the step counts during these activities were very comparable to the StepWatch. The recording of steps during supposedly inactive activities likely happened when participants marked a 15-minute interval in the diary as sedentary but did, in fact, interrupt the sitting time by brief bouts of ambulation or did not start to sit precisely at the start of the 15-minute interval. Overall, with a mean difference of 237 steps per day, the deviation of the devices from the StepWatch was very small during low active and inactive activities, however, and it would consequently hardly affect the correct assessment of someone's PA level in a free-living setting. In contrast, the wrist-worn ActiGraph recorded drastically higher step counts during all low active and inactive activities. This large overestimation of up to >200% probably resulted from gesturing and other types of hand/arm movement that was miscounted as steps despite a disabled LFE, and it would markedly influence the accuracy of PA assessment particularly in sedentary populations. It is interesting that the Garmin Vivofit, albeit likewise wrist-worn, is obviously much less affected by hand movement, which makes this device the much-preferred device for step count assessment in study settings.

It is noteworthy, that unrealistically high step counts were recorded by the StepWatch (criterion measure) compared to all other devices during bouts of cycling. While some steps can be explained by the fact that participants may have walked a few steps during a 15-minute interval that was marked as cycling in the diary, the 125% higher step count of the StepWatch compared to all other devices is likely caused by the fact that the StepWatch miscounts pedaling revolutions as steps, which has been reported before³¹. It is surprising that such miscounting occurs since the acceleration profile during ambulation should differ clearly enough from the much smoother pedaling movement during cycling to be distinguishable. This flaw consequently limits the validity of the device and questions the suitability of the StepWatch to be used to measure steps per day in a free-living setting or as a reference method in future studies, particularly in populations that use bicycles as a common mode of transportations such as many European countries³².

A limitation of the study is the fact that video recording was not used as the criterion measure. While the StepWatch has proven validity to record steps in a free-living setting, a MAPE of around 4%¹¹ has to be factored in when interpreting our results, and the suitability of the StepWatch as a reference method is therefore debatable. The use of video recording would have certainly improved the accuracy of the true step counts; however, due to the 3-day monitoring period, we deemed the use of the StepWatch more feasible and less likely to

affect participants' compliance. Nevertheless, this is a clear limitation of our study and future studies should aim to repeat our procedures with a more adequate reference method such as video recording. A further limitation is that activities of daily living were not objectively measured but instead self-recorded in 15-minute intervals in a diary. Intervals of 15 minutes further have the disadvantage that the recording of brief interruptions of extended sedentary time is not possible; however, shorter time-intervals would have likely influenced participants' compliance in filling out the diary and consequently not led to a gain in accuracy. Finally, the standardized and uninterrupted placement of the smartphones in the pants pocket does not reflect actual use of smartphones in free-living conditions. While previous results in the laboratory⁵ show that the validity of step detection of the iPhone SE is independent of the phone's placement in the pants pocket, a shoulder bag or a backpack, thus justifying the standardized placement in the pants pocket in this study, future studies should assess these devices' accuracy in detecting steps during actual ad libitum everyday use. This would allow a more real-life assessment of their suitability to be used for PA assessment in free-living conditions.

CONCLUSION

In conclusion, the overall high MAPE of all devices, particularly during step-based activities, is likely due to an undercounting of steps during brief, intermittent bouts of ambulation and limits their validity in a free-living setting. However, as shown by the unrealistically high and likely false steps counts that were recorded by the StepWatch during cycling, even the most valid device is not without flaws and the suitability of the StepWatch to detect and measure steps in a free-living setting can therefore not unequivocally be confirmed. When using smartphones, consumer activity wristbands, and even research-grade, hip-worn accelerometers for PA assessment in study settings, a general underestimation of steps per day of around 20% should be factored in when interpreting the results. Particularly in inactive individuals, the Garmin Vivofit should be preferred over the wrist-worn ActiGraph. The Garmin Vivofit is less affected by gesturing and hand movement and consequently records less false steps during sedentary activities of daily living, which allows for a more accurate assessment of daily PA.

Acknowledgments:

The authors are grateful to all participants for their time and effort, which were crucial to the success of this research, and to the ETH Zurich, Switzerland, for providing the StepWatch devices that were used in this study.

Funding: This study was funded by the PhD Program Health Sciences (PPHS) of the University of Basel, Switzerland. During the revision of this manuscript, CH was funded by an NIH NIDDK National Research Service Award (T32 DK064584).

REFERENCES

1. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. Comparison of wrist-worn and hip-worn activity monitors under free living conditions. *J Med Eng Technol.* 2017;41(3):200–207. doi:10.1080/03091902.2016.1271046 [PubMed: 28078908]
2. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med.* 2014;48(13):1019–1023. doi:10.1136/bjsports-2014-093546 [PubMed: 24782483]

3. Kerr J, Marinac CR, Ellis K, et al. Comparison of Accelerometry Methods for Estimating Physical Activity. *Med Sci Sports Exerc.* 2017;49(3):617–624. doi:10.1249/MSS.0000000000001124 [PubMed: 27755355]
4. Statista: Number of smartphone users worldwide 2014–2020. <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>. Published 2017 Accessed March 16, 2017.
5. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. *Scand J Med Sci Sports.* 2018;28(7):1818–1827. doi:10.1111/sms.13074 [PubMed: 29460319]
6. An H-S, Jones GC, Kang S-K, Welk GJ, Lee J-M. How valid are wearable physical activity trackers for measuring steps? *Eur J Sport Sci.* 2017;17(3):360–368. doi:10.1080/17461391.2016.1255261 [PubMed: 27912681]
7. Fokkema T, Kooiman TJM, Krijnen WP, van der Schans CP, de Groot M. Reliability and Validity of Ten Consumer Activity Trackers Depend on Walking Speed. *Med Sci Sports Exerc.* 2017;49(4):793–800. doi:10.1249/MSS.0000000000001146 [PubMed: 28319983]
8. Hekler EB, Buman MP, Grieco L, et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. *JMIR MHealth UHealth.* 2015;3(2):36. doi:10.2196/mhealth.3505
9. Major MJ, Alford M. Validity of the iPhone M7 motion co-processor as a pedometer for able-bodied ambulation. *J Sports Sci.* 2016;34(23):2160–2164. doi:10.1080/02640414.2016.1189086 [PubMed: 27240005]
10. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil.* 2015;7:24. doi:10.1186/s13102-015-0018-5 [PubMed: 26464801]
11. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. *Med Sci Sports Exerc.* 1 2018. doi:10.1249/MSS.0000000000001569
12. Webber SC, Magill SM, Schafer JL, Wilson KCS. GT3X+ accelerometer, Yamax pedometer and SC-StepMX pedometer step count accuracy in community-dwelling older adults. *J Aging Phys Act.* 2014;22(3):334–341. doi:10.1123/japa.2013-0002 [PubMed: 23921227]
13. Barriera TV, Tudor-Locke C, Champagne CM, Broyles ST, Johnson WD, Katzmarzyk PT. Comparison of GT3X accelerometer and YAMAX pedometer steps/day in a free-living sample of overweight and obese adults. *J Phys Act Health.* 2013;10(2):263–270. [PubMed: 22821951]
14. Feng Y, Wong CK, Janeja V, Kuber R, Mentis HM. Comparison of tri-axial accelerometers step-count accuracy in slow walking conditions. *Gait Posture.* 2017;53:11–16. doi:10.1016/j.gaitpost.2016.12.014 [PubMed: 28064084]
15. Chu AHY, Ng SHX, Paknezhad M, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLoS ONE.* 2017;12(2). doi:10.1371/journal.pone.0172535
16. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sci Sports Exerc.* 2015;47(4):839–842. doi:10.1249/MSS.0000000000000476 [PubMed: 25121517]
17. Allen MJ, Yen WM. *Introduction to Measurement Theory.* Waveland Press; 2001.
18. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S531–543. doi:10.1249/01.mss.0000185657.86065.98 [PubMed: 16294116]
19. Ried-Larsen M, Brønd JC, Brage S, et al. Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *Int J Behav Nutr Phys Act.* 2012;9:113. doi:10.1186/1479-5868-9-113 [PubMed: 22971175]
20. Aadland E, Ylvisåker E. Reliability of the Actigraph GT3X+ Accelerometer in Adults under Free-Living Conditions. *PLoS ONE.* 2015;10(8). doi:10.1371/journal.pone.0134606
21. Mudge S, Stott NS, Walt SE. Criterion validity of the StepWatch Activity Monitor as a measure of walking activity in patients after stroke. *Arch Phys Med Rehabil.* 2007;88(12):1710–1715. doi:10.1016/j.apmr.2007.07.039 [PubMed: 18047890]

22. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of Activity Monitor Step Detection Is Related to Movement Patterns. *J Phys Act Health*. 2016;13(2):145–153. doi:10.1123/jpah.2015-0203 [PubMed: 26107045]
23. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sports Med Auckl NZ*. 2017;47(7):1303–1315. doi:10.1007/s40279-016-0663-1
24. ActiGraph LLC. Low Frequency Extension Explained. <https://actigraph.desk.com/customer/portal/articles/2515505>. Published Pensacola, FL 2016 Accessed October 11, 2018.
25. Toth LP, Bassett DR, Crouter SE, et al. StepWatch accuracy during walking, running, and intermittent activities. *Gait Posture*. 2017;52:165–170. doi:10.1016/j.gaitpost.2016.11.035 [PubMed: 27914311]
26. Foster RC, Lanningham-Foster LM, Manohar C, et al. Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Prev Med*. 2005;41(3–4):778–783. doi:10.1016/j.ypmed.2005.07.006 [PubMed: 16125760]
27. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med*. 2012;31(29):3972–3981. doi:10.1002/sim.5466 [PubMed: 22764084]
28. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12:159. doi:10.1186/s12966-015-0314-1 [PubMed: 26684758]
29. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? For adults. *Int J Behav Nutr Phys Act*. 2011;8:79. doi:10.1186/1479-5868-8-79 [PubMed: 21798015]
30. Toth L, Park S, Pittman W, Sarasaltik D, Morton AL, Bassett DRF. Step Count Filters in Wearable Step Counters: 1355 Board #30 June 1 9: 00 AM - 10: 30 AM. *Med Sci Sports Exerc*. 5 2017. doi: 10.1249/01.mss.0000517885.68396.5b
31. Karabulut M, Crouter SE, Bassett DR. Comparison of two waist-mounted and two ankle-mounted electronic pedometers. *Eur J Appl Physiol*. 2005;95(4):335–343. doi:10.1007/s00421-005-0018-3 [PubMed: 16132120]
32. Transportation Research Board. Making Transit Work: Insight from Western Europe, Canada, and the United States -- Special Report 257. Washington, DC: National Academy Press; 2001. doi: 10.17226/10110

Figure 1:
Difference in measured steps of the devices compared to the StepWatch (dotted line).

Figure 2: Bland-Altman plots comparing steps per day using the StepWatch and (A) iPhone SE, (B) Garmin Vivofit_1 (distal wrist), (C) Garmin Vivofit_2 (proximal wrist), (D) ActiGraph wGT3X+_1 (hip), (E) ActiGraph wGT3X+_2 (hip), and (F) ActiGraph wGT3X+_3 (wrist).

Figure 3:
Number of steps recorded by the devices during different activities of daily living.

Figure 4:

Absolute difference between Garmin Vivofit_1 (distal wrist) and Garmin Vivofit_2 (proximal wrist), and Actigraph_1 and Actigraph_2 (both placed at the hip).

Table 1.

Participant characteristics.

Characteristic	N = 30
Female sex, n (%)	18 (60)
Age, yr	25 (23, 32)
Height, cm	170 (164, 178)
Body mass, kg	64 (54, 75)
Body mass index, kg·m ⁻²	22 (20, 24)
Data recording time, hr:mm/day	13:30 (12:18, 14:23)

Data are median (interquartile range) if not stated otherwise.