The validity of consumer-level activity monitors in healthy older adults in free-living conditions

Farina, N., & Lowry, R.G. (2017) The validity of consumer-level activity monitors in healthy older adults in free-living conditions. The Journal of Aging and Physical Activity. https://doi.org/10.1123/japa.2016-0344

Abstract

Consumer-level activity monitors, such as Fitbit and Misfit devices, are a popular and low-cost means of measuring physical activity. This study aims to compare the accuracy of step counts from two consumer-level activity monitors against two reference devices in healthy, community dwelling older adults in free-living conditions. Twenty-five older adults (aged 65-84) simultaneously wore 5 devices (e.g. Misfit Shine and Fitbit Charge HR) over 7 consecutive days. All consumer-level activity monitors positively correlated with reference devices ($p<0.001$). There was also *substantial* to *near perfect* agreement between all consumer-level activity monitors and reference devices. Compared to the Actigraph GT3X+, the waist worn Misfit Shine displayed the highest agreement amongst the devices worn (ICC=0.96, 95% 0.91 to 0.99). The wrist worn devices showed poorer agreement to reference devices. Future research needs to consider that not all consumer-level activity monitors are equal in terms of accuracy, design and function.

Key words: physical activity, accuracy, accelerometer, Fitbit, Misfit, older adults

The validity of consumer-level activity monitors in older adults in free-living conditions

The benefits of physical activity are well established, being able to prevent or delay the onset of a variety of chronic diseases (Singh, 2002), as well as improving physical fitness, reducing depression, anxiety, and mortality (Dunn, Trivedi, & O'Neal, 2001; Hupin et al., 2015; Warburton, Nicol, & Bredin, 2006). In addition, an active lifestyle for older adults is associated with reduced risk of musculoskeletal conditions such as arthritis and osteoporosis (Broskey et al., 2014). Within research and clinical settings, activity monitors are commonly used as a means to objectively and conveniently measure physical activity levels without adding to participant burden.

In the past few years, consumer-level activity monitors have gained popularity, which has led to the widespread availability of accelerometer technology. The range of devices available vary in terms of cost, information recorded, battery life (or frequency of charging), device displays and support applications. A major limitation to the implementation of consumer-level activity monitors in research and clinical settings is the absence of robust scientific evidence that these devices are valid in terms of data captured. In adults, there have been a number of studies that have attempted to validate these devices in lab settings (e.g. Fulk et al., 2014; Kooiman et al., 2015; Lee, Kim, & Welk, 2014). These studies find that a number of consumer-level activity monitors are valid, though the accuracy varies between devices. These devices have also been validated in "free-living" conditions (Ferguson, Rowlands, Olds, & Maher, 2015; Kooiman et al., 2015; Tully, McBride, Heron, & Hunter, 2014), which are likely to better reflect real world activities.

Notably, the validation of these devices in the general adult population are unlikely to be generalizable to older adults. Studies have found gait speed and the use of a walking aid were important factors in determining the accuracy of activity monitors (Paul et al., 2015; Phillips, Petroski, & Markis, 2015) and pedometers (Cyarto, Myers, & Tudor-Locke, 2004). As such, activity monitors are likely to become less accurate as a person ages, with decreased walking speeds, need for walking aids and mobility restrictions that could also compromise posture and movement.

Very few studies have explored the validity of consumer-based activity monitors in older adults (Floegel, Florez-Pregonero, Hekler, & Buman, 2016; Paul et al., 2015; Phillips et al., 2015). Often these studies compared consumer-level activity monitors against visual count and research grade accelerometers in a lab setting. To our knowledge only one of these studies validated such monitors in free-living conditions (Paul et al., 2015). Paul and colleagues did determine the validity of a single device (Fitbit) worn in a single location (waist). However, as previously identified, the validity of consumer-level activity monitors is likely to vary depending on bodily location worn as well as the device type (Floegel et al., 2016). As monitors have evolved, there is increasing variation in the models' design, function, device placement and price. Therefore understanding the validity of multiple consumer activity-monitors, positioned on different bodily locations, is useful in giving clinicians and researchers guidance on the best device to use in older adults.

This study will validate two commercially available devices (Misfit Shine and Fitbit Charge HR) against two well-validated research-grade, tri-axial activity monitors (ActiGraph GT3X+ and New Lifestyle NL2000i) in community-dwelling older adults in free-living conditions. A

secondary aim was to determine whether the Misfit Shine, which is designed to be worn on the wrist or waist, is valid to wear on one or both locations.

Methods

Study population

Participants were community-dwelling older adults (aged 65-84), recruited from West Sussex, England and Co. Down, Northern Ireland. Participants were excluded if they were not independently ambulatory or used a walking aid (self-reported). There was no exclusion criteria stipulated for the amount of physical activity that participants were currently undertaking.

Procedure

The xxxxxx Research Ethics Committee approved this study and all participants provided informed consent prior to commencing the study.

Following informed consent, participants were given a series of questionnaires. Participants were then asked to wear five activity monitors on different locations on their body (i.e. wrist and waist). The three waist worn devices were attached to an elastic belt and positioned above the dominant kneecap. The two wrist worn monitors were positioned on the dominant wrist, the Fitbit Charge HR located nearest the hand and the Misfit Shine immediately above. All five devices were adjusted for the participant's gender, age, height, weight and handedness prior to being worn. Participants were given a series of walking activities on a number of terrains to assess the validity of the measures in a lab-setting (methods and results not described here).

Following the initial visit, participants were asked to continue to wear the devices for a full week and return to the lab at the same time the following week. Participants were instructed to put the monitors on and take them off at the same time and to wear the activity monitors during waking hours, except during bathing and water-based sports. Participants were not asked to change their daily habits during the study, and were not given access to device software or informed of additional device features. Participants were instructed not to interfere with the device (except to charge the Fitbit device via a USB connector) or change the device location. On their return visit to the lab, participants completed additional questionnaires including their self-reported physical activity habits over the previous week. Compliance of wearing the devices in accordance with the protocol was assessed through verbal confirmation from participant of adherence upon returning the devices, and visual confirmation that the devices were correctly positioned upon returning the devices.

Measures

Demographic information was collected (i.e., age, gender, ethnicity, handedness, years of education, number of falls in the past year, and subjective physical complaints). Participants' height and weight were also measured and body mass index was calculated.

International Physical Activity Questionnaire (IPAQ; Craig et al., 2003) is a 27 item questionnaire on physical activity participation over the past 7 days. It includes questions about activities at work, house-work, getting from place to place, as well as for leisure and sport. The IPAQ has an adequate validity in older adults (Tomioka, Iwamoto, Saeki, & Okamoto, 2011). Vigorous, moderate and low intensity activity is accumulated and metabolic equivalent time (METs) can be calculated.

Charlson Comorbidity Index (CCI; Charlson, Pompei, Ales, & MacKenzie, 1987) is a 19 item measure of comorbidity. Each comorbidity category has an associated weight (from 1 to 6), based on the adjusted risk of mortality or resource use, and the sum of all the weights results in a single comorbidity score.

Montreal Cognitive Assessment (MOCA; Nasreddine et al., 2005) is a short measure of global cognitive impairment. Higher scores represent better cognitive performance.

Short Physical Performance Battery (SPPB; Guralnik et al., 1994), a measure of functional performance through measures of standing balance, strength and endurance. Higher scores represent better functional performance.

Reference Devices

NL2000i (*New-Lifestyles Inc, Lees Summit, Missouri, USA*) activity monitor is a research grade, tri-axial (3 dimensional), piezoelectric accelerometer. It provides daily data in the form of total steps, active steps, active (moderate to vigorous physical activity, MVPA) minutes, distance, active calories and total calories expended. The device has been found to demonstrate small differences in total steps recorded to those observed (+ 3 steps) over a 400 meter distance (Schneider, Crouter, Lukajic, & Bassett, 2003), and energy expenditure against Doubly Labelled Water (DLW) method (Colbert, Matthews, Havighurst, Kim, & Schoeller, 2011). The NL-2000i memory capacity is 7 days plus the current day which tallies rest at 00:00 hours. This data can be accessed from a digital display on the device. The device was worn on the waistband of clothing above the midpoint (the kneecap) on the dominant side. The NL2000i was programmed with the time and the individual's age, gender, height and weight.

ActiGraph GT3X+ (*Actigraph, Pensacola, Florida, USA*) is an accelerometer that records tri-axial movement. Actigraph devices are commonly used accelerometers for measuring physical activity in research studies (Chen & Bassett, 2005). The device continually records sedentary time, activity (light, moderate and vigorous) behaviour, estimates energy expenditure and counts steps. The Actigraph GT3X reference device has a good agreement of steps against visually observed step count (McMinn, Acharya, Rowe, Gray, & Allan, 2013) and energy expenditure against DLW (Van Remoortel et al., 2012). The device was worn on the waist, above the midpoint on the dominant side. The Actigraph was programmed with the time, device location and the individual's age, gender, height, weight and hand dominance; all were set to record movement in 10-second epochs. Upon return to the lab, the data was extracted from the device using the Actilife software package. The Normal Filter (digital filtering algorithm) was applied, which is designed to eliminate noise outside of the human activity frequency bandwidth (Cain, Conway, Adams, Husak, & Sallis, 2013).

Consumer-level activity monitors

The Fitbit Charge HR *(Fitbit, Inc., San Francisco, California, USA)* is a tri-axial motion sensor, which records activity in 60-second epochs. It provides daily data in the form of steps, distance, calories, activity intensity and sleep. This brand of device has previously been used in research with an older population (Paul et al., 2015; Phillips et al., 2015). It requires charging via a USB cable every three to four days (depending on usage). The device was set up with participant's gender, age, height, weight and hand dominance. This data can be accessed from a digital display on the device or by synchronizing the device to a website or smartphone application. The device was worn on the wrist on the dominant side.

The Misfit Shine *(Misfit Wearables, Burlingame, California, USA)* motion sensor records steps, distance, calories burned, sleep quality and duration and active time. This device was selected, in part, because of its low-profile design and longevity of battery. It does not provide a digital display on the device and therefore has to be synchronized with the accompanying application to obtain the data. The device was worn on the wrist using the strap accessory and on the elastic waistband with the Actigraph and the NL2000i on the dominant side. The device was set up with participant's gender, age, height, weight and hand dominance.

Analysis Plan

The sample size of this study was chosen to be in line with previous activity monitor validation studies in free-living conditions (Ferguson et al., 2015; Kooiman et al., 2015). Based upon a systematic review of the literature step count of consumer-level activity monitors in free-living conditions correlated with reference devices above r=0.80 (Evenson, Goto, & Furberg, 2015). With α =0.05 and β=0.20 a sample size of 9 participants would be needed to detect whether a correlation of 0.80 differs from zero (Hulley, 2007).

A total score for the CCI, SPPB and MOCA were created and the mean score (and standard deviation) for each were reported to provide a description of health characteristics. For the IPAQ, the median total MET-min/week (and interquartile range) was calculated based on the scoring guidance (IPAQ Research Committee, 2005). Cases were excluded from the total IPAQ score and associated subdomain if they contained any missing data in the relevant time or day's responses or had unreasonably high minutes' worth of activity (i.e. > 960 minutes/day).

An average steps/day for each device was calculated for every participant. This was calculated by summing the total number of steps taken between testing and dividing by the number of valid whole days (i.e. 6 days). In line with previously guidance (Colley, Connor Gorber, & Tremblay, 2010), we implemented quality control and data reduction procedures on device data to ensure compliance. A cut-off threshold of 150 minutes of continuous zero data from the Actigraph was deemed as being 'non-wear' data (Hutto et al., 2013). Similar to previous research (Troiano et al., 2008), 10 hours per day of Actigraph data was required for data to be considered valid. If either criteria was not met, data from that day (across all devices) were removed. A minimum of 4 days of valid data were required for inclusion in analysis, in line with common practice (Tudor-Locke, Camhi, & Troiano, 2012).

Normality was assessed (Shapiro-Wilk Test of Normality) on the step count data for each activity monitor to determine the use of non-parametric or parametric techniques. Initially a series of Spearman's rank correlations were used to compare steps/day between consumer-level devices and reference devices. Bonferroni-Holm's correction for multiple tests (Holm, 1979) was performed and only corrected P values less than 0.05 were considered to be significant. Intraclass correlation coefficients (ICC 2,1) were then used to compare the agreement between step counts taken from the consumer-level activity monitors and the two reference devices (ActiGraph GT3X+ and NL2000i). Previous conservative guidelines have suggested that ICCs <0.75 are poor to moderate, ≥ 0.75 are good, and >0.90 are reasonable for "clinical measurements" (Portney & Watkins, 2015). Bland-Altman plots were used to visualise any systematic differences between step counts from the consumer-level activity monitors and the two reference devices.

A series of Spearman's rank correlations were also used to compare the amount of time participants spent physically active in the past week (based on the IPAQ) compared to the number of steps/day for each device. Bonferroni-Holm's correction for multiple tests (Holm,

1979) was performed and only corrected P values less than 0.05 were considered to be significant.

Data was analysed using SPSS V.23 (*IBM Corporation, Armonk, New York, USA*).

Results

A total of 25 participants consented to be involved in the study. Approximately half of participants were female (n=12). The sample were relatively, physically healthy. Only two participants reported having a physical complaint, neither of whom felt it affected their day-to-day physical activity. Participants also scored highly on the SPPB, an indicator of good physical fitness and stability. Demographic and health characteristics are displayed in Table 1. Of the 25 participants that wore the activity monitors, 9 had missing data from at least 1 activity monitor. Reasons for missing data included: loss of device, researcher error in setup of device, and repositioning of device.

A single participants' daily step count was classed as an extreme outlier (Misfit Shine – waist; 153,012 steps in a single day) and removed. Across all participants, 4 participants had evidence of at least one day of non-wear Actigraph data, whilst 7 participants had at least one day of Actigraph data less than 10 hours. Three participants' data were excluded from analysis because they did not have at least 4 days of valid Actigraph data. The total score for the IPAQ could not be calculated in 9 cases; 6 participants had missing data in the relevant time or days responses and 3 participants had unreasonably high minutes' worth of activity.

Based upon the Actigraph reference device, on average the participants took 7503.67 (SD= 3526.26) steps/day. Comparatively, the NL2000i reference device recorded an average of 8350.42 (SD=3906.67) steps/day. The two reference devices had near perfect agreement (ICC = 0.96, 95% CI: 0.89 to 0.98).

Misfit Shine (Waist)

On average, the waist worn Misfit Shine (waist) device underestimated the number of steps/day compared to the NL2000i device and overestimated compared to the Actigraph device (Table 2). The Misfit Shine positively correlated with both the Actigraph and NL2000i reference devices. The Misfit Shine also had near perfect agreement compared to the Actigraph and the NL2000i. The relationship is depicted in the Bland-Altman plots (Figure 1).

Misfit Shine (Wrist)

The wrist worn Misfit Shine (wrist) also underestimated the number of steps/day compared to both reference devices. The Misfit Shine (wrist) positively correlated with both the Actigraph and NL2000i reference devices. Compared to the Actigraph and NL2000i, the Misfit Shine (Wrist) had good agreement, though with wide CIs. The limits of agreement were moderately wide. See Table 2 and Figure 1.

Fitbit Charge HR (Wrist)

The wrist worn Fitbit Charge HR had good agreement compared to the Actigraph and the NL2000i, positively correlating with both. The wrist worn Fitbit Charge HR did however display systematic bias, substantially overestimating the step count compared to both the Actigraph and NL2000i. The limits of agreement were also very wide against the Actigraph and NL2000i. See Table 2 and Figure 1.

Self-reported weekly physical activity

The total physically active time spent in the past week, based upon the IPAQ, did not significantly correlate with steps/day as measured by activity monitors (See Table 3).

Discussion

This study set out to validate two commercially available activity monitors in a healthy older adult population in free-living conditions, whilst determining whether the placement of such devices impact their accuracy.

As expected, all consumer-level activity monitors strongly and positively correlated with both reference devices. In line with the findings from a sample of healthy adults (BodyMedia SenseWear; ICC =0.90) (Ferguson et al., 2015), the wrist worn Misfit Shine showed excellent agreement compared to reference devices. Interestingly, it was found that whilst the Misfit Shine showed excellent agreement on both the wrist and the waist, the agreement was much lower for the wrist worn device. This could be attributed to the fact that the reference devices were also placed on the waist, and therefore are more likely to pick up similar movement. Waist located placement of such devices is often preferable due to their close proximity to the central mass of the body, and thus is likely to represent major movement (Yang & Hsu, 2010).

The second consumer-level activity monitor, the Fitbit Charge HR, had excellent agreement compared to both reference devices but displayed wide confidence intervals and limits of agreement. The average steps/day measured over the testing period was much greater than the number of reference steps, indicating that there is systematic bias. Fitbit devices (Fitbit One and Zip) have previously been validated in older adults (Paul et al., 2015). Whilst it is not clear whether the underlying technologies within the Fitbit family are the same, it is possible

that the variation in accuracy is as a result of device placement (i.e. The Fitbit Charge HR - wrist worn, Fitbit One and Zip - waist worn).

In validating these activity monitors, it is important to consider the sample characteristics as the findings may not generalizable to other populations. In the present study, all participants were ambulatory, physically healthy older adults. There is an indication that physical activity levels in the present sample is in line with the wider older adult population, with steps/day (M=7143.6, Actigraph) and MET-min/week (Mdn=5343.0, IPAQ) being similar to those reported elsewhere (Lohne-Seiler, Hansen, Kolle, & Anderssen, 2014; Tudor-Locke et al., 2011). Cognitively, this sample displayed some evidence of impairment, with participants on average scoring below cut-offs (MOCA<26)(Davis et al., 2015), though this is in line with age-related population norms (Rossetti, Lacritz, Cullum, & Weiner, 2011).

A key limitation of this study is that the accuracy of the consumer-level devices are based upon the agreement with existing reference devices, and therefore is on the assumption that in an older population, these reference devices are accurately capturing all physical activity. As previously highlighted, daily activities in free-living conditions are likely to involve a substantial amount of upper body movement which may not be captured by typical waist worn activity monitors (Lee et al., 2014). The fact that the Misfit devices, regardless where positioned, did not significantly overestimate the number of steps compared to waist worn reference devices suggest that this is unlikely to undermine its accuracy. It is however plausible that device placement has greater importance in older adults that use walking aids (Floegel et al., 2016). Another limitation of this study is that the population was composed of healthy older adults who were independently ambulatory and therefore did not use a walking aid; scoring near ceiling on the SPPB and a low CCI score supports this. As a result, the findings reported here may not reflect those from a physically impaired older adult population, with gait speed and use of walking aid affecting the accuracy of activity monitors

(Paul et al., 2015; Phillips et al., 2015). The final limitation of this study is that we did not have any objective means to determine whether participants wore the device in accordance with the protocol. However, we utilised quality control and data reduction to decrease the likelihood of noncompliance.

Gaining a better understanding of the validity of consumer-level activity monitors in older adults is an important first step in researchers and clinicians adopting them. The increasing evidence base highlights that a number of commercially available devices can accurately measure physical activity levels in older adults is promising. However, until we understand if the specifications within a brands activity monitors are similar, the validity and reliability of new devices to the market need to be confirmed (Evenson et al., 2015). Ultimately, it will be a near impossible task to validate every device that comes to the market. Efforts to validate a range of devices, with different design features, will provide researchers and clinicians options in selecting a device that best suits their objectives and target population.

Conflict of Interest

There is no known conflict of interest. The authors declare that results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

References

Broskey, N. T., Greggio, C., Boss, A., Boutant, M., Dwyer, A., Schlueter, L., … Amati, F. (2014). Skeletal muscle mitochondria in the elderly: effects of physical fitness and exercise training. *The Journal of Clinical Endocrinology and Metabolism*, *99*(5), 1852–1861. https://doi.org/10.1210/jc.2013-3983

Cain, K. L., Conway, T. L., Adams, M. A., Husak, L. E., & Sallis, J. F. (2013). Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. *International Journal of Behavioral Nutrition and Physical Activity*, *10*, 51. https://doi.org/10.1186/1479-5868-10-51

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383. https://doi.org/10.1016/0021-9681(87)90171-8

Chen, K. Y., & Bassett, D. R. (2005). The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Medicine & Science in Sports & Exercise*, *37*(Supplement), S490–S500. https://doi.org/10.1249/01.mss.0000185571.49104.82

Colbert, L. H., Matthews, C. E., Havighurst, T. C., Kim, K., & Schoeller, D. A. (2011). Comparative validity of physical activity measures in older adults. *Medicine and Science in Sports and Exercise*, *43*(5), 867.

Colley, R., Connor Gorber, S., & Tremblay, M. S. (2010). Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health Reports*, *21*(1), 63–69.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., … Oja, P. (2003). International Physical Activity Questionnaire: 12-Country

Reliability and Validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Cyarto, E. V., Myers, A. M., & Tudor-Locke, C. (2004). Pedometer accuracy in nursing home and community-dwelling older adults. *Medicine and Science in Sports and Exercise*, *36*(2), 205–209. https://doi.org/10.1249/01.MSS.0000113476.62469.98

Davis, D. H., Creavin, S. T., Yip, J. L., Noel-Storr, A. H., Brayne, C., & Cullum, S. (2015). Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010775.pub2/abstract

Dunn, A. L., Trivedi, M. H., & O'Neal, H. A. (2001). Physical activity dose-response effects on outcomes of depression and anxiety: *Medicine and Science in Sports and Exercise*, *33*(Supplement), S587–S597. https://doi.org/10.1097/00005768-200106001-00027

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, *12*, 159. https://doi.org/10.1186/s12966-015-0314-1

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 42. https://doi.org/10.1186/s12966-015-0201-9

Floegel, T. A., Florez-Pregonero, A., Hekler, E. B., & Buman, M. P. (2016). Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, glw098. https://doi.org/10.1093/gerona/glw098

Fulk, G. D., Combs, S. A., Danks, K. A., Nirider, C. D., Raja, B., & Reisman, D. S. (2014). Accuracy of 2 Activity Monitors in Detecting Steps in People With Stroke and Traumatic Brain Injury. *Physical Therapy*, *94*(2), 222–229. https://doi.org/10.2522/ptj.20120525

Guralnik, J. M., Simonsick, E. M., Ferrucci, L., Glynn, R. J., Berkman, L. F., Blazer, D. G., … Wallace, R. B. (1994). A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology*, *49*(2), M85-94.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Hulley, S. B. (2007). *Designing Clinical Research* (3rd Edition). Philadelphia, USA: Lippincott Williams & Wilkins.

Hupin, D., Roche, F., Gremeaux, V., Chatard, J.-C., Oriol, M., Gaspoz, J.-M., … Edouard, P. (2015). Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: a systematic review and meta-analysis. *British Journal of Sports Medicine*, bjsports-2014-094306. https://doi.org/10.1136/bjsports-2014-094306

Hutto, B., Howard, V. J., Blair, S. N., Colabianchi, N., Vena, J. E., Rhodes, D., & Hooker, S. P. (2013). Identifying accelerometer nonwear and wear time in older adults. *International Journal of Behavioral Nutrition and Physical Activity*, *10*, 120. https://doi.org/10.1186/1479-5868-10-120

IPAQ Research Committee. (2005). Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)–short and long forms. Retrieved from

https://www.researchgate.net/post/Can_anyone_explain_the_rules_of_Data_truncation_for_the_IPAQ_long_form

Kooiman, T. J. M., Dontje, M. L., Sprenger, S. R., Krijnen, W. P., van der Schans, C. P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, *7*. https://doi.org/10.1186/s13102-015-0018-5

Lee, J.-M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1840–1848. https://doi.org/10.1249/MSS.0000000000000287

Lohne-Seiler, H., Hansen, B. H., Kolle, E., & Anderssen, S. A. (2014). Accelerometer-determined physical activity and self-reported health in a population of older adults (65–85 years): a cross-sectional study. *BMC Public Health*, *14*, 284. https://doi.org/10.1186/1471-2458-14-284

McMinn, D., Acharya, R., Rowe, D. A., Gray, S. R., & Allan, J. L. (2013). Measuring activity energy expenditure: accuracy of the GT3X+ and actiheart monitors. *International Journal of Exercise Science*, *6*(3), 217–229.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Paul, S. S., Tiedemann, A., Hassett, L. M., Ramsay, E., Kirkham, C., Chagpar, S., & Sherrington, C. (2015). Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. *BMJ Open Sport & Exercise Medicine*, *1*(1), e000013. https://doi.org/10.1136/bmjsem-2015-000013

Phillips, L. J., Petroski, G. F., & Markis, N. E. (2015). A Comparison of Accelerometer Accuracy in Older Adults. *Research in Gerontological Nursing*, *8*(5), 213–219. https://doi.org/10.3928/19404921-20150429-03

Portney, L. G., & Watkins, M. P. (2015). *Foundations of clinical research: applications to practice* (3rd ed.). Philadelphia, USA: F.A. Davis Company. Retrieved from https://www.atsu.edu/policies/documents/3%20ashs%20-%202013-14%20athletic%20training%20textbook%20list.pdf

Rossetti, H. C., Lacritz, L. H., Cullum, C. M., & Weiner, M. F. (2011). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*, *77*(13), 1272–1275. https://doi.org/10.1212/WNL.0b013e318230208a

Schneider, P. L., Crouter, S. E., Lukajic, O., & Bassett, D. R. (2003). Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Medicine and Science in Sports and Exercise*, *35*(10), 1779–1784. https://doi.org/10.1249/01.MSS.0000089342.96098.C4

Singh, M. A. F. (2002). Exercise Comes of Age Rationale and Recommendations for a Geriatric Exercise Prescription. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *57*(5), M262–M282. https://doi.org/10.1093/gerona/57.5.M262

Tomioka, K., Iwamoto, J., Saeki, K., & Okamoto, N. (2011). Reliability and Validity of the International Physical Activity Questionnaire (IPAQ) in Elderly Adults: The Fujiwara-kyo Study. *Journal of Epidemiology*, *21*(6), 459–465. https://doi.org/10.2188/jea.JE20110003

Troiano, R. P., Berrigan, D., Dodd, K. W., MâSse, L. C., Tilert, T., & Mcdowell, M. (2008). Physical Activity in the United States Measured by Accelerometer: *Medicine &*

Science in Sports & Exercise, *40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Tudor-Locke, C., Camhi, S. M., & Troiano, R. P. (2012). A Catalog of Rules, Variables, and Definitions Applied to Accelerometer Data in the National Health and Nutrition Examination Survey, 2003–2006. *Preventing Chronic Disease*, *9*. https://doi.org/10.5888/pcd9.110332

Tudor-Locke, C., Craig, C. L., Aoyagi, Y., Bell, R. C., Croteau, K. A., Bourdeaudhuij, I. D., … Blair, S. N. (2011). How many steps/day are enough? For older adults and special populations. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 1–19. https://doi.org/10.1186/1479-5868-8-80

Tully, M. A., McBride, C., Heron, L., & Hunter, R. F. (2014). The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Research Notes*, *7*, 952. https://doi.org/10.1186/1756-0500-7-952

Van Remoortel, H., Raste, Y., Louvaris, Z., Giavedoni, S., Burtin, C., Langer, D., … others. (2012). Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PloS One*, *7*(6), e39198.

Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006). Health benefits of physical activity: the evidence. *CMAJ : Canadian Medical Association Journal*, *174*(6), 801–809. https://doi.org/10.1503/cmaj.051351

Yang, C.-C., & Hsu, Y.-L. (2010). A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, *10*(8), 7772–7788. https://doi.org/10.3390/s100807772

Table 1. *Demographics and health characteristics of sample (n=25)*

	N (Valid Percentage)	Mean (SD)
Age		72.5 (4.9)
Gender	Male: 13 (52.0%)	
	Female: 12 (48.0%)	
Residential Location	City: 5 (20.0%)	
	Town: 6 (24.0%)	
	Village: 14 (56.0%)	
Ethnicity/Nationality	White British: 24 (96.0%)	
	White Irish: 1 (4.0%)	
Handedness	Left: 3 (12.0%)	
	Right: 14 (88.0%)	
Falls in the past year	Yes: 4 (16.7%)	
	No: 20 (83.3%)	
Falls resulting in hospitalisation	Yes: 0 (0.0%)	
	No: 4 (16.7%)	
	N/A: 20 (83.3%)	
Any physical complaints	Yes: 2 (8.7%)	
	No: 21 (91.3%)	
BMI		27.4 (3.8)
Age leaving full-time education (years)		16.7 (2.2)
MOCA (Max = 30)		24.3 (3.1)
CCI total score		3.0 (0.8)
SPPB total score (Max = 12)		11.7 (0.9)
Device wear time (mins/day)		843.1 (65.3)
		Median (IQR)
IPAQ (MET-min/week)		5343.00 (4016.75)

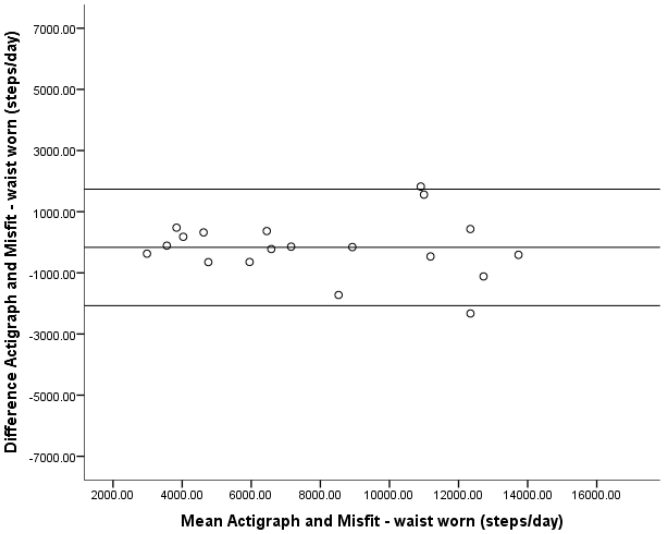
BMI= Body Mass Index; CCI= Charlson Comorbidity Index; MET = metabolic equivalent of task; MOCA= Montreal Cognitive Assessment; SPPB= Short Physical Performance Battery; SD=Standard Deviation; IQR=Interquartile Range; IPAQ = International Physical Activity Questionnaire

Table 2. *Consumer-level activity monitors steps/day compared to reference devices (Actigraph GT3X+ and NL2000i)*

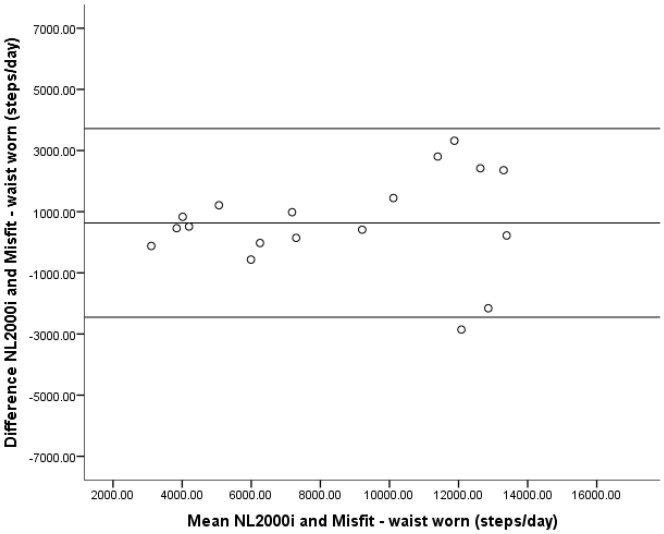
Reference Device		Misfit Shine -Waist	Misfit Shine - Wrist	Fitbit Charge HR- Wrist
Actigraph GT3X+[a]				
	Mean Difference (SD)[c]	-167.6 (972.1)	10.1 (1690.7)	-2690.3 (2014.3)
	r_s	0.96*	0.91*	0.84*
	ICC 2,1 (95% CI)	0.96 (0.91 to 0.99)	0.86 (0.67 to 0.94)	0.86 (0.68 to 0.94)
	Limits of Agreement	-2072.9 to 1737.7	-3303.7 to 3323.9	-6638.3 to 1257.7
NL2000i[b]				
	Mean Difference (SD)[c]	633.2 (1574.7)	899.7 (2014.4)	-1721.6 (2225.7)
	r_s	0.90*	0.89*	0.83*
	ICC 2,1 (95% CI)	0.91 (0.79 to 0.97)	0.83 (0.59 to 0.93)	0.85 (0.65 to 0.94)
	Limits of Agreement	-2453.2 to 3719.6	-3048.5 to 4847.9	-6084.0 to 2640.8

[a]Mean (SD) = 7503.7 (3526.3); [b] Mean (SD)= 8350.4 (3906.7); [c] positive values indicate underestimation, negative values indicate overestimation. *p<0.001

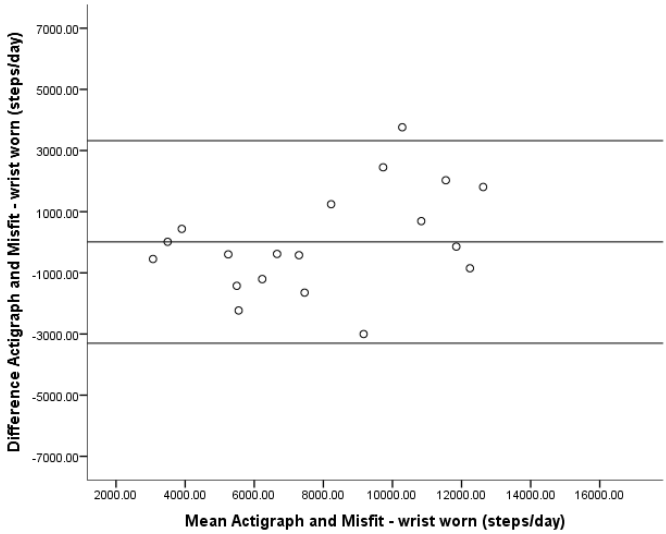
A


Difference Actigraph and Misfit - waist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean Actigraph and Misfit - waist worn (steps/day)

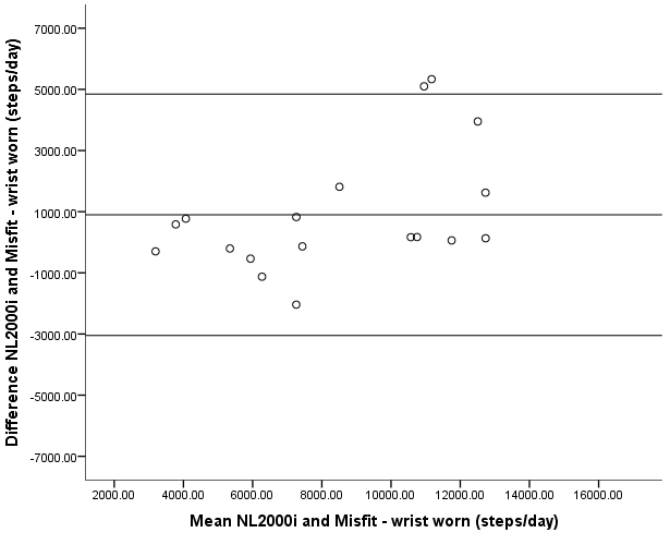
B


Difference NL2000i and Misfit - waist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean NL2000i and Misfit - waist worn (steps/day)

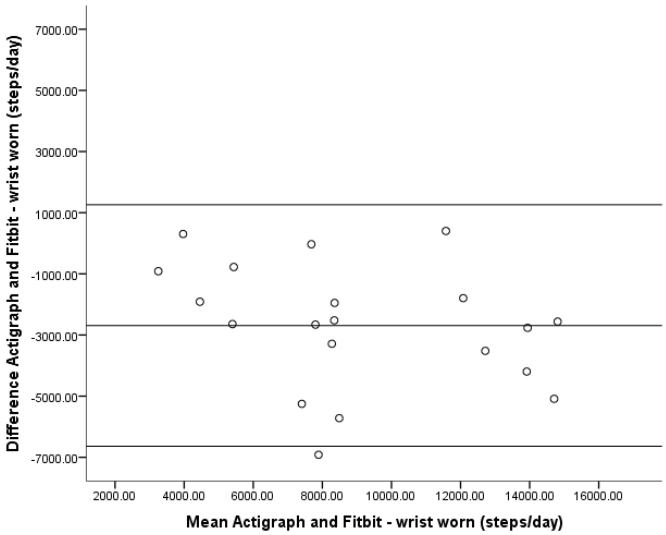
C


Difference Actigraph and Misfit - wrist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean Actigraph and Misfit - wrist worn (steps/day)

D


Difference NL2000i and Misfit - wrist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean NL2000i and Misfit - wrist worn (steps/day)

E


Difference Actigraph and Fitbit - wrist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean Actigraph and Fitbit - wrist worn (steps/day)

F

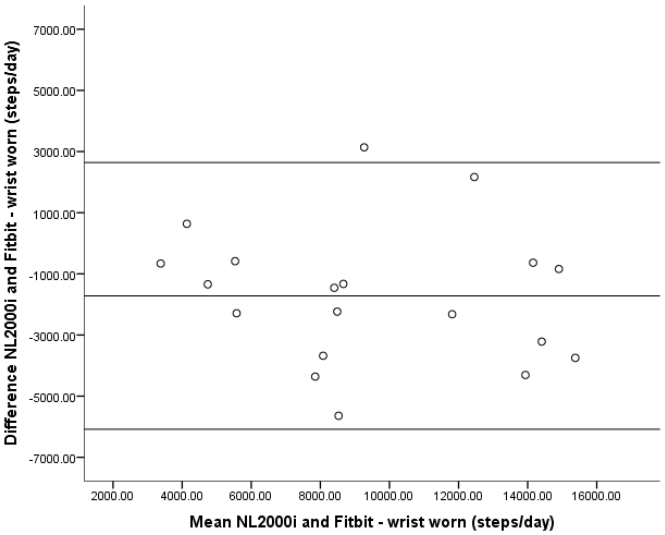

Difference NL2000i and Fitbit - wrist worn (steps/day)
7000.00
5000.00
3000.00
1000.00
-1000.00
-3000.00
-5000.00
-7000.00
2000.00
4000.00
6000.00
8000.00
10000.00
12000.00
14000.00
16000.00
Mean NL2000i and Fitbit - wrist worn (steps/day)

Table 3. *Spearman correlation coefficients between activity monitors (steps/day) and time per week spent exercising based on the IPAQ.*

	Consumer-level devices			Reference devices	
	Misfit Waist (Steps/day)	Misfit Wrist (Steps/day)	Fitbit Wrist (Steps/day)	NL2000i (Steps/day)	Actigraph GT3X+ (Steps/day)
IPAQ Total Physical Activity (mins/week)	-0.12	-0.02	0.11	-0.14	-0.06

IPAQ= International Physical Activity Questionnaire.