Degroote, L1,2,5, MSc; Hamerlinck, G1; Poels, K3, PhD; Maher, C4, PhD; Crombez, G5, PhD; De Bourdeaudhuij, I1, PhD; Vandendriessche, A6, MSc; Curtis, RG4, PhD; DeSmet, A1,2,3,7, PhD

1 Ghent University, Department of Movement and Sports Sciences, Belgium
2 Research Foundation Flanders, Belgium
3 University of Antwerp, Department of Communication Studies, Belgium
4 University of South Australia, School of Health Sciences, Alliance for Research in Exercise, Nutrition and Activity (ARENA), Australia
5 Ghent University, Department of Experimental Clinical and Health Psychology, Belgium
6 Ghent University, Department of Public Health and Primary Care, Belgium
7 Université Libre de Bruxelles, Belgium

Low-cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: A Validation Study

Abstract

Background: Wearable trackers for monitoring physical activity (PA) and total sleep time (TST) are increasingly popular. These devices are not only used by consumers to monitor their behavior, but also by researchers to track the behavior of large samples and health professionals to implement interventions aimed at health promotion and to remotely monitor patients. However, high cost and accuracy concerns may be barriers to widespread adoption.

Objective: This study investigated the concurrent validity of six low-cost activity trackers: Geonaut On Coach, iWown i5 Plus, MyKronoz ZeFit4, Nokia GO, VeryFit 2.0 and Xiaomi MiBand 2 for measuring steps, moderate to vigorous physical activity (MVPA) and total sleep time (TST).

Methods: A free-living protocol was used in which 20 adults engaged in their usual daily activities and sleep. For 3 days and 3 nights, they simultaneously wore 1 low-cost tracker and 1 high-cost tracker (Fitbit Charge HR) on the non-dominant wrist. Participants wore an ActiGraph GT3X+ accelerometer on the hip at daytime and wore a BodyMedia SenseWear device on the nondominant upper arm at nighttime. Validity was assessed by comparing each tracker with the ActiGraph GT3X+/BodyMedia
SenseWear using Mean Absolute Percentage Error scores (MAPE), correlations and Bland-Altman plots in IBM SPSS 24.0.

**Results:** Large variations were shown between trackers. Low-cost trackers showed moderate to very strong correlations (Spearman r=0.53-0.91) and low to good agreement (interclass correlation coefficient, ICC=0.51-0.90) for measuring steps. Weak to moderate correlations (Spearman r=0.24-0.56) and low agreement (ICC=0.18-0.56) were shown for measuring MVPA. For measuring TST, the low-cost trackers showed very weak to strong correlations (Spearman r=0.04-0.73) and low agreement (interclass correlation coefficient, ICC=0.05-0.52). Bland-Altman revealed a variation between over- and undercounting for measuring steps, MVPA and TST depending on the used low-cost tracker. None of the trackers, including the Fibit (high-cost one), showed high validity to measure MVPA.

**Conclusions:** This study was the first to examine the concurrent validity of low-cost trackers. Validity was strongest for measurement of steps, there was evidence of validity for measurement of sleep in some trackers, whereas validity for measurement of MVPA time was weak throughout all devices. Validity ranged between devices, with the Xiaomi having the highest validity for measurement of steps, and the VeryFit performing relatively strong across both sleep and steps domains. Low-cost trackers hold promise for monitoring and measurement of movement and sleep behaviors, both for consumers and researchers.

**Keywords:** low-cost, activity tracker, accelerometry, concurrent validity, steps physical activity, total sleep time
Introduction

Physical activity (PA) and sleep are modifiable determinants of morbidity and mortality among adults, and specifically contribute to development of diseases such as obesity, type 2 diabetes, cardiovascular diseases, low quality of life and mental health problems [1-6]. Getting at least 30 minutes moderate-to-vigorous physical activity (MVPA) per day, getting between 7-9 hours in total sleep time (TST) per night, and spending relatively more time on light physical activity (LPA) rather than being sedentary, are associated with beneficial health outcomes [1-6]. A large part of adults does not meet the guidelines for one or more of these behaviors [6, 7]. PA and sleep are, together with time spent on sedentary behavior (SB), co-dependent behaviors: they are part of one 24-hour day and time spent on one behavior will impact the time spent on at least one of the other behaviors. It is therefore recommended to target these behaviors together [8].

Successful health promotion interventions rely on behavior change techniques that address modifiable determinants of health behavior [9]. A behavior change technique reported as both effective [10] and highly appreciated by users [11, 12], is self-monitoring of health behavior. Self-monitoring refers to keeping a record of the behavior that is performed [13]. Self-monitoring tools provide opportunities for self-management of health, as well as for remote activity tracking by health care providers as part of a patient’s treatment regimen [14]. Subjective ways of self-monitoring, such as self-report, retrospective measures, often face high participant burden and reporting biases and errors, leading to substantial overestimations of the healthy behavior [14]. Self-reported sleep duration in sleep logs showed an overestimation in comparison to objective measurements, especially when sleep duration was below recommended health norms [15]. Activity trackers conversely offer an automated objective and convenient mean for self-monitoring PA and sleep. This paper focuses on self-monitoring via consumer-based activity trackers as intervention tools for PA and sleep, more specifically by investigating the validity of low-cost trackers. Such trackers rarely monitor SB (see [16, 17]), which is why SB, albeit important in 24-hour movement behaviors, falls outside of the scope of this paper.

Activity trackers may include pedometers (‘step counters’), smartphone-based accelerometers, and accelerometers in advanced electronic wearable trackers or in smartwatches. Pedometers, however, do not provide information on sleep, and smartphone-based accelerometers have shown lower accuracy to measure PA compared to advanced electronic wearable trackers [18], making advanced electronic wearable trackers and smartwatches more suited to accurately self-monitor PA and sleep. Smartwatches (e.g. Apple Watch) offer several other functions apart from activity tracking such as communication and entertainment, and are usually more expensive than advanced electronic wearable trackers (e.g. Fitbit Charge). Advanced electronic wearable trackers (termed as activity trackers hereafter) are usually wrist- or belt-worn, provide 24-hour self-monitoring and often include real-time behavioral feedback or more detailed feedback shown after synchronization with other electronic devices (such as tablet, smartphone or PC) [19]. Several commercial
activity trackers are available to the public, and are increasingly integrated in effective intervention programs to improve activity behaviors [20, 21].

There has been an increased interest by adults in activity trackers. For example, in Flanders (Belgium), 8% of adults had an activity tracker (22% all wearables included, also sport- and smartwatches) in 2018 compared to only 2% (8% all wearables included, also sport- and smartwatches) owning one in 2015 [22]. Characteristics of activity trackers may impact their continued use and further adoption. Cost is likely to be a barrier for increased adoption of higher-end trackers [19, 23]. Indeed, activity trackers appear to be used less among adults who are lower-educated, unemployed [19], and have a lower income [22]. Notably, unhealthy lifestyles such as insufficient PA [24] and insufficient sleep duration [25] are more prevalent among people of lower to medium socio-economic status (SES) than among those of higher SES. So providing accurate low-cost options to self-monitor PA and sleep in their daily lives is crucial for public health as a lack of valid low-cost trackers may increase the health and digital divide between lower and higher income groups in society. However, non-adoption of activity trackers in low SES populations can probably not only be attributed to the high cost of the devices, but may also be a matter of priorities and affordances. Further research in this area is necessary. Having valid low-cost trackers not only plays a role in low SES populations: also in the general population, cost-effective solutions are needed for scaling-up interventions in a public health context where financial resources are limited [26]. Having accurate low-cost activity trackers can be expected to increase the feasibility of scaling-up interventions that rely on activity trackers.

The unequal access to valid tools due to cost barriers are often studied within health literacy conceptual frameworks. Health literacy refers to having the ability and motivation to take responsibility for one’s own health [27]. Low health literacy has been associated with worse health outcomes [27] and improving having access to tools that can help understand their own health behaviour via self-monitoring and take responsibility to take care of one’s own health, may improve health literacy. There is increasing attention to expanding the Health Literacy model to eHealth literacy or digital health literacy, defined as the ability of people to use emerging technological tools to improve or enable health and health care [28]. Digital health literacy appears associated with a lower socio-economic status [29].

More specifically, the importance of accuracy of low cost trackers can also be understood from the Technology Acceptance Model that emphasizes the need for trust and perceived usefulness, together with perceived ease of use, of a tool prior to users being willing to adopt them [30].

When using activity trackers, their accuracy needs to be established to avoid counterproductive effects, such as falsely signaling that people are meeting guidelines and need not make any extra efforts whereas in fact these people may not reach sufficient sleep or PA [31]. Conversely, an underestimation of actual behavior can also cause people to get demotivated and to no longer make efforts to do better [32].
Accuracy of the tracker has also been cited by users as trackers’ most important characteristic [19]. To effectively use wearable activity trackers for health self-management in daily life, accuracy needs to be assessed in free-living settings, since laboratory-based validity studies tend to over-estimate validity [18]. The validity to measure physical activity in free-living conditions has been examined for several activity trackers, such as the Fitbit One, Zip, Ultra, Classic, Flex [16-18, 33-38], Misfit Shine [18] and Withings Pulse [18]. In general, studies found the highest validity for Fitbit trackers [18]. Most validated trackers showed high correlations with an ActiGraph accelerometer for number of steps [16, 39, 40]. MVPA is less often studied and less accurately measured by activity trackers than step counts [39]. Activity trackers showed moderate-to-strong correlations with ActiGraph accelerometers on MVPA, with Fitbit trackers and Withings Pulse showing highest accuracy [39]. Also for TST, several wearable activity trackers currently on the market have been assessed for validity, including Fitbit (Flex and Charge HR) [41-43], Withings Pulse [39, 41], Basis Health Tracker [41], Garmin [44], and Polar Loop [44]. Validity results for TST were very divergent, ranging from low to strong validity, with Fitbit again showing better validity [39, 44]. Accuracy of PA and/or TST depends on the position where the tracker is worn, e.g. wrist versus hip [33] and can be improved by combining accelerometry with heart rate measurement [45, 46].

The cost of the trackers in above-mentioned published validation studies was often not reported, but their price in the current market (based on Amazon.de at the end of January 2019) ranged from €50 to €130 for an unused, basic model (with Misfit Flash as the exception, at €42). Most trackers that are popular in the consumer market and that are reported on in scientific publications cost more than €50, and commonly more than €100 [14]. A recent industry report states that when spending less than USD$50, users are likely to get a product of mediocre accuracy [47], though it is unclear whether this statement was empirically based. To our knowledge, only two studies have examined the validity of low-cost trackers. Wahl and colleagues’ study of the Polar Loop (price June 2019 around €60 on Amazon.de), Beurer AS80 (price June 2019 around €42) and Xiaomi Mi Band (price June 2019 around €25) suggested that only the Mi Band had good validity for step count [48]. However, this study was conducted in laboratory, and not in free-living conditions. In one other study, validity of the Xiaomi Mi Band for measuring TST was evaluated relative to manual switch-to-sleep-mode measurement, with positive results [49]. This study, however, did not use an objective measurement tool as comparison. We are not aware of any validation studies of low-cost activity trackers against objective measurement methods conducted in free-living conditions, and many of the most commonly available low-cost trackers do not appear to have been validated in any form.

In summary, wearable activity trackers can be a useful tool in health promotion and remote treatment monitoring for PA and TST. However, high cost and accuracy concerns may be barriers to widespread adoption [50]. Assessing validity of low-cost trackers may play a major role at population level to encourage health behavior in the future, also in low SES groups who are most at risk for poor health and in need of healthy behavior promotion. To enable activity self-monitoring in daily life, the
accuracy of low-cost wearable activity trackers needs to be established in free-living conditions. Current validation studies have mainly focused on wearable activity trackers that cost above €50. This current study aims to assess the validity of low-cost wearable activity trackers among adults (≤€50), for the objective measurement of PA and TST in daily life against free-living gold standards (ActiGraph GT3X+ accelerometer and BodyMedia SenseWear). This study is exploratory in nature and does not have firm hypotheses regarding the validity of specific low-cost trackers. However, it may be expected that trackers with a heart-rate monitoring are more accurate than those without. This, because they may provide more robust estimates of intensity and energy expenditure used to discriminate between activity and nonactivity [45, 46].

Methods

A concurrent validity study among adults was designed in which a low-cost tracker was validated against a free-living-condition standard for steps, active minutes (MVPA) and for TST. A high-cost tracker (Fitbit Charge 2) was also validated against these gold standards, to compare with validation outcomes for the low-cost trackers. In each participant, three 24-hour observation days were collected for each low-cost tracker. Power analyses (run in G*Power 3.1.9.2) suggested that to detect a two-tailed significant correlation $H_1 = 0.49 - 0.90$, with 80% power (values based on Brooke et al., 2017), a sample size of between $n=6$ and $n=29$ was required.

Participants and procedure

Twenty healthy participants between 18 and 65 years of age living in Flanders, Belgium, were recruited using convenience sampling. Inclusion criteria were having no current physical limitations, medical conditions, or psychiatric conditions that may impact movement or sleep. Descriptive information collected on participants consisted of age, sex, self-reported height and weight, and highest attained education. All participants read and signed an informed consent form. The study protocol was approved by the Ethics Committee of the University hospital of Ghent (B670201731732).

Instruments

Convergent Measure
As this is a free-living study, the ActiGraph GT3X+ (Actigraph, Pensicola, FL, USA) tri-axial accelerometer was used as a reliable and valid reference for measuring step counts [51-53], and MVPA [54, 55]. The GT3X+ has been shown to be a valid measure of both step count compared with direct observation (percentage error <1.5%[52]; percentage error ≤1.1% [53]; ICC ≥ 0.84 [56]) and MVPA compared to indirect calorimetry ($r = 0.88$) [54]. Accelerometer data were initialized, downloaded and processed using ActiLife version 5.5.5-software (ActiGraph, Fort Walton Beach, FL, USA). The Freedson Adult (1998) cut-points were applied to categorize physical
activity measured by the ActiGraph accelerometer (sedentary activity = 0–99 counts/min, light activity = 100–1951 counts/min, moderate activity = 1952–5723 counts/min, and vigorous activity = ≥5724 counts/min) [57]. A 15 sec epoch was used when downloading the data. The ActiGraph GT3X+ was fitted to the right side of the participants’ waist in accordance with the manufacturer’s instructions. Only days with valid data of the ActiGraph were included in the analysis. A valid day was defined as a 24-hour period in which at least 10 hours of data wear time was recorded [58]. Non-wear time was analysed as a run of zero counts lasting more than 60 min with an allowance of 2 min of interruptions. Using this algorithm, the risk of misclassification of non-wear time as sedentary time was avoided [59].

The BodyMedia SenseWear (BodyMedia Inc., Pittsburg, PA, USA) is a portable multisensor device that can provide information regarding the total energy expenditure, TST, circadian rhythm and other activity metrics. In this study, the SenseWear was used as the references for sleep duration. SenseWear has been validated as a measure of TST compared with polysomnography (r=0.83, standard error of estimate=37.71) [60]. Data were analysed in SenseWear Professional 8.1 software [61]. The SenseWear was placed over the triceps muscle on the non-dominant arm between the acromion and olecranon processes, in accordance with the manufacturer’s instructions.

**Low-cost activity trackers**

Six low-cost activity trackers were selected (Figure 1) based on their price at the time of the study (≤€50); their market share (e.g. MyKronoz, Xiaomi); whether or not they included a heart-rate measurement, output (steps; MVPA or active minutes; TST); and availability from popular online purchase sites in Europe, where the study was conducted. Furthermore, we tested the Fitbit Charge 2 to also include a comparison between a low-cost and a validated high-cost activity tracker. Fitbit was selected as a high-cost activity tracker because it was one of the most popular activity trackers on the market at the start of the study (IDC, 2018) and was already validated for measuring steps, MVPA and TST [17]. All participants received a Wiko smartphone in loan (Lenny 3, Android 6.0 Marshmallow, price €99,99 in June 2019) to pair the trackers with, to cancel out any potential individual differences in smartphone pairing.
Figure 1. Tracker characteristics

All devices measured steps and TST. Only the Xiaomi, Nokia and also the Fitbit used a specific variable that quantifies intensive forms of physical activity. These three devices, reported ‘Active Minutes’ with no further subdivision. As the devices all set a goal of 30 minutes physical activity per day (similar to the MVPA recommendations for adults), it was assumed that the measured variable corresponded to MVPA as measured by the ActiGraph. However, specific information regarding intensity cut-points is not publically available. TST was used, excluding daytime naps, for comparison with the SenseWear that was only worn at night. Only the Fitbit, VeryFit and Xiaomi measured heart rate. Data were extracted using the proprietary software for all devices, in the same fashion that a consumer would utilise the software, and were visually checked for outliers.

Free-living Protocol

As it was not feasible nor comfortable to wear all trackers at the same time, participants were instructed to wear one of the low-cost devices in combination with the Fitbit on their non-dominant wrist. They were also instructed to simultaneously
wear the ActiGraph on their hip during daytime and the SenseWear on their upper arm at night-time. Furthermore the participants were provided with a diary to write down the time they put on and took off the devices. This way it could be checked that the devices were always worn simultaneously. If this was not the case, data of the device that was worn separately, was deleted in order to avoid a mismatch of the measurements. Participants received the six low-cost trackers in a random order. The position of low-cost and high-cost tracker on the non-dominant wrist (1st or 2nd in distance from wrist) was varied across days. Each tracker was worn for a period of three consecutive days and nights. A period of three days and nights was chosen to balance between achieving sufficient data for the question under study without burdening the participants. Between two periods, a one-day gap allowed for switching the devices. During daytime, the devices were worn during all waking hours, except during water-based activities. When participants went to bed they were asked to remove the ActiGraph and put on the SenseWear instead. In Figure 2, a typical measurement period for one device is shown.

Figure 2. Example of measurement protocol for one period

PA or TST may differ between weekdays and weekend days. Although this study did not intend to explain differences in PA or TST, but rather the degree of agreement between two measurements on any given day, a difference in how often a tracker was measured on a certain day rather than another day may influence validity results. For example, validity has shown to be lower for measuring a very low number of steps or high number of steps. Our study design controlled for this potential influence by randomly varying the days across participants on which a particular tracker was worn. Across all data points we would then expect all measurement days to relatively equally represented, as was the case in our study. The percentage of weekend days in total measurement days ranged between 25% and 33%. Also on particular weekdays, there were very few differences (2-9% difference between the tracker with the lowest number of measurements on a certain day and the tracker with the highest number of measurements on a particular day).
**Statistical analysis**

Analyses were performed using IBM SPSS Statistics version 24.0 (SPSS Inc., Chicago, IL, USA). All analyses were performed on a daily measurement level, counting a measured day as unit of analysis. Analyses consisted of measures of agreement, systematic differences, and bias and limits of agreement. Measures of agreement (equivalence testing) included: 1) Spearman correlation coefficient r to examine the association between steps, active minutes and TST measured by trackers and convergent measure (also illustrated in scatter plots). As sleep and physical activity data was non-normally distributed, the Spearman correlation, a non-parametric statistical test, was used instead of Pearson correlation; and 2) Intraclass Correlation Coefficient ICC (absolute agreement, two-way random, single measures, 95% confidence interval) that reflects the effect of individual differences on observed measures. Measures of systematic differences included mean absolute percentage errors (MAPE) of tracker measurements compared to that of the convergent measure. MAPE were calculated with the following formula: mean difference activity tracker-convergent measure x 100 /mean gold standard. Bland-Altman plots with their associated limits of agreement examined biases between measurements from the trackers and the convergent measure. The following cut-off values were used to interpret the Spearman r: <0.20 = very weak; 0.20-0.39 = weak; 0.40-0.59 = moderate; 0.60-0.79 = strong; 0.80-1.0 = very strong [62]. The cut-off values to interpret the ICC were: <0.60 = low; 0.60-0.75 = moderate; 0.75-0.90 = good; >0.90 = excellent [63].

A series of linear mixed effects models with restricted maximum likelihood estimation examined the association between steps, MVPA minutes and TST measured by the commercial trackers and convergent measures, accounting for the structure of the data (repeated measures clustered within participants). The pattern of results was similar to those obtained by the above analyses. Data are therefore presented in Supplementary Table 1.

**Results**

**Descriptive Statistics**

Three participants discontinued their participation to the study: 1 dropped out at the start of the study due to the combination of high perceived burden of the research protocol and a busy personal schedule, consequently no data was collected and analysed from this participant; 1 was not able to meet the protocol towards the end of the study due to conflict with his/her work schedule; 1 had to end participation due to an unexpected hospital admission (retention rate 17/20, 85%). The average age of the analysed sample of participants who started the study (n=19) was M=37.6 ±13.4; 13/19 were female. The sample was highly educated, with 17/19 having
achieved a higher education degree (academic or non-academic). Their average Body Mass Index was 23.5 ± 4.4. Two participants were overweight (BMI 25-30), two were obese (BMI ≥30). The level of MVPA measured at baseline with the International Physical Activity Questionnaire (IPAQ) varied from 10 to 351 minutes per day (SD=91)[64, 65].

All participants owned a smartphone; 5 out of 19 participants had previous experience with wearable trackers (n=3 Fitbit). As can be expected in a higher educated sample, they were all very familiar with digital tools and required little assistance in installation or usage. We do not expect any impact of participants’ experience on the validity measurements, as 1) these would not have a differential effect of any potential misuse between different trackers, and 2) control procedures were put in place to prevent any misuse. Potential misuse could consist of a wrong placement of the tracker. Participants received a thorough briefing at the start of the study and a daily check-up of any issues, to ensure any baseline differences in familiarity with digital tools were cancelled out and to reduce the risk for misuse. No issues with misuse were noted.

**Issue of usability with low-cost trackers**

In total each device was intended to be tested for 60 days. As one of the participants did not start, the maximum number of potential measurement days per tracker was reduced to 57. The number of days of available data varied per tracker: 1) due to drop-out at the end of the study for some participants; 2) due to technical issues experienced with some trackers, which resulted in fewer days of available data.

The VeryFit had 55/57 (96%) measured days for PA (lost days: n=2 no data shown in app) and 51/57 (86%) for sleep (lost days: n=3 participant non-compliance, n=3 no data shown). iWown had 52/57 (89%) measured days for PA (lost days: n=4 tracker did not pair, n=1 no data shown) and 51/57 (89%) for sleep (lost days: n=4 tracker did not pair, n=2 no data shown). Xiaomi was not worn by two participants due to drop-out, reducing potential measurement days to 51. Xiaomi had 48/51 (94%) measured days for PA (lost days: n=2 participant non-compliance, n=1 no data shown) and 44/51 (86%) measured days for sleep (lost days: n=6 participant non-compliance, n=1 no data shown). Nokia had 49/57 (86%) measured days for both PA (8 lost days due to no data shown) and 46/57 for sleep (81%) (lost days: n=8 due to no data shown, n=3 due to participant non-compliance). MyKronoz was not worn by three participants due to drop-out; one participant accidentally removed the data, reducing potential measurement days to 45. MyKronoz had 40/45 (89%) measurement days for PA (lost days: n=5 no data shown in app), and 24/45 (53%) for sleep (lost days: n=11 no data shown in app, n=10 participant non-compliance). Geonaut had 37/57 (65%) measured days for PA (lost days: n=12 no data shown, n=9 did not pair, n=5 participant non-compliance), and 30/57 (53%) measured days for sleep (lost days: n=9 did not pair, n=8 no data shown, n=4 participant non-compliance).

Participants were especially frustrated about a device not pairing, as this meant they had to reinstall the tracker and also lost the history of their past activity. In sum,
VeryFit and Xiaomi showed very little data loss due to usability problems; whereas especially for Geonaut and MyKronoz data were lost due to usability problems. In general, more data were lost for sleep than for PA. Usable data in the analyses were further reduced due to technical issues experienced with the convergent measures, which resulted in fewer days of data for which comparisons could be made (usable data shown in Tables 1, 3).

**Validity of low-cost trackers**

**Physical Activity**

Table 1 shows the mean steps, mean minutes of MVPA and the corresponding standard deviations for all trackers for measuring steps and MVPA.

Table 1. Mean steps and minutes of MVPA per day measured by the low-cost trackers, the Fitbit and the ActiGraph

| Tracker       | Number of measured days | Mean ± SD      | Minimum | Maximum |
|---------------|-------------------------|----------------|---------|---------|
| **Number of steps per day** |                         |                |         |         |
| Geonaut       | 37                      | 8026 ± 4352    | 657     | 19413   |
| iWown         | 51                      | 7668 ± 5169    | 259     | 22759   |
| MyKronoz      | 40                      | 10431 ± 4764   | 485     | 24493   |
| Nokia         | 50                      | 5896 ± 3113    | 325     | 13976   |
| VeryFit       | 55                      | 7320 ± 4481    | 649     | 22628   |
| Xiaomi        | 48                      | 7317 ± 4535    | 369     | 20866   |
| Fitbit        | 307                     | 9662 ± 4866    | 451     | 24664   |
| ActiGraph     | 316                     | 8126 ± 4314    | 188     | 23121   |

| **Number of minutes of MVPA** |                     |       |                   |         |
|------------------------------|---------------------|-------|-------------------|---------|
| Nokia                        | 49                  | 5 ± 12| 0                  | 52      |
| Xiaomi                       | 46                  | 80 ± 48| 0                  | 190     |
| Fitbit                       | 305                 | 45 ± 49| 0                  | 239     |
| ActiGraph                    | 328                 | 41 ± 31| 0                  | 150     |

a MVPA: moderate-to-vigorous physical activity.

Agreement testing for steps diverged between Spearman r and ICC (Table 2). All trackers, except the iWown, showed strong (Nokia, Geonaut, VeryFit and MyKronoz) to very strong (Xiaomi and Fitbit) agreement with the ActiGraph measurements based on the Spearman r (all above 0.60). Based on the ICC, the MyKronoz, iWown and Nokia showed low agreement (ICC<0.60), whereas the Geonaut had moderate and the Xiaomi, Fitbit and VeryFit had a good agreement with the ActiGraph measurements (ICC 0.75-0.90). These coefficients are in line with the interpretation of the MAPE scores, showing the largest mean deviation from the ActiGraph measurements for the iWown (35.28%) and the smallest for the Xiaomi tracker (17.14%).
For measuring MVPA, correlations between the MVPA measurements of the trackers and the ActiGraph accelerometer were weak for the Nokia and the Xiaomi and moderate for the Fitbit (Table 2). The ICC showed low agreement for MVPA between all three trackers and the ActiGraph accelerometer (ICC<0.60). The MAPE scores also indicate very large mean deviations from the ActiGraph measurements for MVPA (>100%), which confirm the low accuracy of the trackers for measuring MVPA.

Table 2. Correlation coefficients, intraclass correlation coefficients, associated 95% CI of the measurements and MAPE scores for measuring steps and MVPA.

| Tracker | N  | Spearman r | 95% CI | ICC  | 95% CI | MAPE (%) |
|---------|----|-------------|--------|------|--------|----------|
| **Steps** | | | | | | |
| Geonaut | 36 | 0.63<sup>a</sup> | 0.31-0.87 | 0.68<sup>a</sup> | 0.46-0.82 | 24.63 |
| iWown  | 50 | 0.53<sup>a</sup> | 0.16-0.77 | 0.51<sup>a</sup> | 0.28-0.69 | 35.28 |
| MyKronoz | 38 | 0.77<sup>a</sup> | 0.45-0.95 | 0.59<sup>a</sup> | 0.22-0.79 | 25.79 |
| Nokia  | 50 | 0.77<sup>a</sup> | 0.51-0.94 | 0.56<sup>a</sup> | 0.27-0.74 | 22.62 |
| VeryFit | 54 | 0.78<sup>a</sup> | 0.61-0.89 | 0.82<sup>a</sup> | 0.62-0.91 | 24.87 |
| Xiaomi  | 45 | 0.91<sup>a</sup> | 0.81-0.97 | 0.90<sup>a</sup> | 0.77-0.95 | 17.14 |
| Fitbit | 300 | 0.91<sup>a</sup> | 0.86-0.94 | 0.87<sup>a</sup> | 0.66-0.93 | 25.73 |
| **MVPA**<sup>b</sup> | | | | | | |
| Nokia  | 16 | 0.24 | -0.11-0.50 | 0.18 | -0.10 ; 0.44 | 108.17 |
| Xiaomi  | 45 | 0.26 | -0.08-0.54 | 0.15 | -0.08 ; 0.39 | 293.29 |
| Fitbit | 298 | 0.56<sup>a</sup> | 0.47-0.63 | 0.56<sup>a</sup> | 0.48 ; 0.64 | 114.30 |

<sup>a</sup>P<.001.

<sup>b</sup>MVPA: moderate-to-vigorous physical activity.

Correlations for steps and MVPA are illustrated in Figure 3 and Figure 4. Scatter and deviation of the points around the line that reflects the perfect agreement between the measurements, is larger for measuring MVPA than for measuring steps. The largest scatter for measuring steps is found for the iWown (Figure 3). Based on the scatterplots, a careful statement on over- or underestimation of the measurement of the trackers can be made. This is based on the location of the data points relative to the line that represents the perfect agreement between the measurements. For the Xiaomi, Nokia and VeryFit, the majority of the data points is located below that line, meaning an underestimation of the amount of steps. For the iWown, MyKronoz and Fitbit, the majority of the data points are located above the line, meaning an overestimation of the amount of steps. For the Geonaut, no clear under- or overestimation is visualised. These findings are also visualised by the Bland-Altman plots. A large scatter for all three trackers that measures MVPA was observed, with no obvious relation between the MVPA measurements of the trackers and the MVPA measurements of the ActiGraph. For the Nokia, an underestimation is visualised, for the Xiaomi however, an overestimation is visualised. For the Fitbit, no clear under- or overestimation is visualised.
Figure 3. Correlations between steps estimates per day from the trackers and the ActiGraph.
Figure 4. Correlations between MVPA estimates per day from the trackers and the ActiGraph.

Bland-Altman plots visualize the differences between the steps and MVPA measurements of the ActiGraph accelerometer and each tracker (y-axis) against the average number of steps or number of minutes of MVPA of the measurements of these two devices (x-axis). Mean differences with the ActiGraph accelerometer and the limits of agreement are presented in Table 3 (illustrated in Figure 5 and 6 for respectively steps and MVPA). A positive value of the mean difference indicates an underestimation of the measurements of the tracker compared to the ActiGraph measurements, whereas a negative value indicates an overestimation. The systematic over- or underestimation (mean differences) and the range between the upper and lower limits of agreement reflect the accuracy of the measurements of the tracker compared to the measurements of the ActiGraph accelerometer. The broader the range between the lower and the upper limit, the less accurate measurements are.

Table 3. Mean differences of activity measures with the ActiGraph accelerometer and limits of agreement of the activity trackers.

|       | N   | Mean difference of steps (ActiGraph – Tracker) | Limits of Agreement | Range |
|-------|-----|-------------------------------------------------|---------------------|-------|
|       |     |                                                 | Lower              | Upper |
| Steps |     |                                                 |                     |       |
| Geonaut | 36  | -146                                            | -4802              | 4509  | 9311 |
| Iwown  | 50  | 638                                             | -8993              | 10270 | 19263 |
| MyKronoz | 38  | -1798                                           | -5563              | 1967  | 7530 |
|       |     |       |       |       |       |
|-------|-----|-------|-------|-------|-------|
| Nokia | 50  | 1609  | -4229 | 7447  | 11676 |
| VeryFit | 54  | 1356  | -3276 | 5989  | 9265  |
| Xiaomi | 45  | 1011  | -2713 | 4737  | 7450  |
| Fitbit | 300 | -1369 | -5238 | 2499  | 7737  |
| **MVPA**\(^a\) |       |       |       |       |       |
| Nokia | 16  | 32.55 | -18.35| 83.45 | 101.80|
| Xiaomi | 45  | -35.14| -138.96| 68.68 | 207.64|
| Fitbit | 298 | -1.27 | -77.07| 74.52 | 151.59|

\(^a\) MVPA: moderate-to-vigorous physical activity.

For measuring steps and MVPA the table and the plots (Figure 5 and 6) all showed large limits. The Xiaomi tracker showed the narrowest limits (7,450 steps) for measuring steps whereas the iWown showed the broadest limits (19,263 steps). These results are in line with the interpretations of validity findings based on the Spearman r, the ICC and the MAPE score.
Figure 5. Bland-Altman plots of the trackers for measuring steps. The middle line shows the mean difference (Positive values indicate an underestimation of the wearable and negative values indicate an overestimation) between the measurements of steps of the wearables and the ActiGraph, and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).

For MVPA, the ranges between the lower and upper limit of agreement are very large, indicating a low accurate measurement by all three trackers measuring MVPA. The
Bland-Altman plots showed the broadest limits for Xiaomi (207.64 min) and the narrowest limits for Nokia (101.80 min).

Figure 6. Bland-Altman plots of the trackers for measuring MVPA. The middle line shows the mean difference (Positive values indicate an underestimation of the wearable and negative values indicate an overestimation) between the measurements of MVPA of the wearables and the ActiGraph, and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).

In sum, several but not all low-cost trackers showed high accuracy to measure steps. Xiaomi trackers even outperformed the Fitbit tracker in measuring steps. None of the trackers, however, showed good accuracy to measure MVPA, including the Fitbit, that did nevertheless reach a slightly higher validity than the low-cost trackers in measuring MVPA.

**Total Sleep Time (TST)**

Table 4 reports mean minutes of TST and corresponding standard deviations for all trackers.

Table 4. Mean TST per day measured by the low-cost trackers, the Fitbit and the SenseWear.
Spearman correlations between the TST measurements of the trackers and the TST measurements of the SenseWear armband show large diversity between trackers, ranging from very weak (Geonaut) to strong (VeryFit). The ICCs however indicate low agreement (ICC<0.60) between the measurements of all trackers and the measurements of the SenseWear. This could reflect a systematic under- or overestimation of TST by the trackers, which is not evident from Spearman r. The MAPE scores of all trackers also indicate a large mean deviation from the SenseWear measurements for TST, ranging from 20.57% for the Fitbit to 39.08% for the Xiaomi. The correlation coefficients, ICC values, associated 95% CI and MAPE scores for measuring TST are shown in Table 5.

Table 5. Correlation coefficients, intraclass correlation coefficients, associated 95% CI of the measurements and MAPE scores for measuring TST.

| Tracker   | N  | Spearman r | 95% CI    | ICC  | 95% CI    | MAPE(%) |
|-----------|----|------------|-----------|------|-----------|---------|
| TST       |    |            |           |      |           |         |
| Geonaut   | 15 | 0.04       | -0.45-0.60| 0.05 | -0.44-0.52| 26.59   |
| iWown     | 24 | 0.57^a     | 0.19-0.84 | 0.52^b| 0.18-0.76 | 21.33   |
| MyKronoz  | 14 | 0.45       | -0.22-0.86| 0.40^b| -0.07-0.74| 38.15   |
| Nokia     | 24 | 0.66^a     | 0.30-0.88 | 0.30^b| -0.10-0.63| 38.63   |
| VeryFit   | 24 | 0.73^a     | 0.48-0.83 | 0.26 | -0.11-0.61| 30.73   |
| Xiaomi    | 21 | 0.21       | -0.34-0.68| 0.13 | -0.13-0.45| 39.08   |
| Fitbit    | 134| 0.57^a     | 0.40-0.69 | 0.46^a| 0.28;0.60 | 20.57   |

^a P<.001.
^b P<.05.

The correlations for TST are also illustrated in Figure 7. This figure visualizes the large discrepancy between the Spearman correlation coefficient and the ICC, specifically evident for the Nokia and the VeryFit. Although a clear relation is visible between the measurements (Spearman r), almost all data points are above the line that represents the perfect agreement between the measurements. This indicates a systematic overestimation of the TST measurements of the Nokia and the VeryFit.
compared to the convergent measure. Figure 7 also shows the largest scatter for the MyKronoz.

Figure 7. Correlations between TST estimates from the trackers and the SenseWear.

Bland-Altman plots for TST revealed the smallest limits for the VeryFit (263.39) and the broadest limits for the Geonaut (558.25 min). These results are in line with the findings based on the Spearman r and the scatter of the data points. Mean differences
with the SenseWear armband measurements and the limits of agreement are presented in Table 6 and illustrated in Figure 8.

Table 6. Mean differences of TST measures with the SenseWear and limits of agreement of the activity trackers.

|    | N  | Mean difference of TST (Sensewear – Smartwatch) | Limits of Agreement | Range |
|----|----|--------------------------------------------------|---------------------|-------|
|    |    |                                                  | Lower   | Upper |       |
| Geonaut | 15  | 44.93                                            | -234.19 | 324.06 | 558.25 |
| iWown   | 24  | -36.79                                           | -221.22 | 147.63 | 368.85 |
| MyKronoz| 14  | -82.29                                           | -330.55 | 165.98 | 496.53 |
| Nokia   | 24  | -106.46                                          | -293.72 | 80.80  | 374.52 |
| VeryFit | 24  | -97.63                                           | -229.32 | 34.07  | 263.39 |
| Xiaomi  | 21  | -112.14                                          | -355.40 | 131.12 | 486.52 |
| Fitbit  | 134 | -36.91                                           | -213.98 | 140.16 | 354.14 |
Figure 8. Bland-Altman plots of the trackers for measuring TST. The middle line shows the mean difference (Positive values indicate an underestimation of the wearable and negative values indicate an overestimation) between the measurements of TST of the wearables and the SenseWear, and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).

In sum, low-cost trackers showed low (e.g. Geonaut, Xiaomi) to strong (VeryFit) correlations to measure TST, with some trackers such as VeryFit and Nokia
systematically overestimating TST. Fitbit shows low (based on ICC) to moderate (based on Spearman r) validity to measure TST, and is outperformed by VeryFit to measure TST on all indicators of accuracy.

Discussion

This study examined the validity of low-cost trackers (≤€50) for measuring adults’ steps, moderate-to-vigorous physical activity (MVPA) and total sleep time (TST) in free-living conditions. In general, the low-cost trackers were most accurate in the measurement of steps, somewhat accurate for measurement of sleep, and lacked validity for measurement of MVPA time. Validity ranged widely between the various low-cost trackers tested. The performance of the best of the low-cost trackers approached or even exceeded that of the Fitbit Charge 2 (the high-cost comparison tracker), but the worst had weak validity. Notably, the VeryFit 2.0 performed relatively strongly across both sleep and steps domains, however the Xiaomi Mi Band 2 appeared to have the highest validity for measurement of steps.

The finding that many of the low-cost trackers are accurate for measuring steps is promising, given that steps is the metric reported by users of trackers as being of most interest [66]. We found that the low-cost trackers were most accurate for measuring steps in comparison to sleep and minutes of MVPA. This order for validity (i.e. steps > sleep > MVPA) is consistent with findings for these metrics in high-cost trackers [39], though in our study, the low-cost trackers demonstrated weak to moderate validity for MVPA minutes (Spearman’s rho ranged from 0.24 to 0.56) whereas previous research in high-cost trackers has suggested moderate to strong validity (e.g. Ferguson et al study of high-cost trackers reported Pearson’s r ranging from 0.52 to 0.91 [39]). It is possible that some of the differences between the reference values for MVPA derived from the ActiGraph accelerometers and the values recorded by the low-cost trackers, may have originated from measurement error associated with the reference device. Furthermore, a possible explanation for the weak to nil validity found in our study could be that the physical activity variables measured by the low-cost trackers were not explicitly identified as MVPA. However, because all devices had set a goal of 30 minutes physical activity per day (similar to the MVPA recommendations for adults) we assumed that the measured variable corresponded to MVPA as measured by the ActiGraph accelerometer. Nevertheless, specific information regarding algorithm intensity cut-points was not provided and publicly available from these low-cost trackers. Therefore, the discrepancies in this study may be a result of both definitional and measurement problems (e.g. sensitivity algorithm). In this respect it may be very useful in the future, when manufacturers provide more insight into the cut-points and algorithms that were used to translate the raw data into useful information (such as steps and minutes of MVPA).

Whilst research grade accelerometers are the closest we have to a “gold standard” for measurement of MVPA in free-living conditions, the MVPA values derived from them can vary by an order on magnitude depending on parameters such as epoch length
and cut-points [67]. Furthermore, wear position has an impact on validity of MPVA. Studies comparing the validity research-grade accelerometers in different body locations consistently show that the hip position is more accurate than the wrist [68]. Despite the recognised superior validity of hip-worn accelerometers and trackers, over the past 5 or so years, there has been a shift for both consumer trackers and research-grade accelerometers to increasingly be designed for wrist wear, presumably due to improved logistics, such as comfort and convenience. This clear shift in the market highlights that validity should not be considered the be-all and end –all. Issues such as usability, compliance and adherence are also important, even though they tend to receive less attention in the scientific literature.

Evidence for validity of the low-cost trackers for measurement of sleep duration was mixed. Some trackers performed quite strongly. For example, the top performing tracker, the VeryFit 2.0, demonstrated Spearman’s rho of 0.73 for total sleep time compared with the reference device (SenseWear), which was actually superior to the Fitbit Charge HR (rho = 0.57). However, the Bland-Altman analyses revealed the VeryFit 2.0 tended to over-estimate sleep by around 1.5 hours per night compared with the reference device. If this over-estimation were consistent, it could be argued that the data might still be useful for self-monitoring changes in sleep over time. However, the Bland-Altman 95% limits of agreement spanned a range of 263 minutes, suggesting that the extent of over-estimation varied considerable on different administrations. It therefore seems questionable whether the sleep estimates derived from the VeryFit 2.0 are accurate enough to help a user meaningfully monitor/change their sleeping patterns.

The finding that low cost-trackers have strong validity for measuring steps and some validity for measuring sleep is likely to be of interest to public health researchers and clinicians alike. There is considerably interest in using activity trackers to intervene on lifestyle activities, with a recent meta-analysis finding positive evidence for short-term effectiveness, but less evidence for sustained effects [69]. There is well-recognised usage attrition associated with activity trackers over time – e.g. a 2017 study gave entry-level Fitbits to n=711 users, and found that approximately 50% of participants had stopped using them at 6 months, and 80% had stopped by 10 months [70]. The most common reasons for not using the Fitbit was technical failure or difficulty (57%), losing the device (13%) or forgetting to wear it (13%). Nonetheless, low-cost devices fill an important gap in the consumer market – between the high-cost activity trackers which are prohibitively expensive to provide to clinical or research cohorts at scale (unless sizable funding is available) but likely to be more aesthetically-pleasing and acceptable to wearers than traditional pedometers. [19, 66]. The findings of this study, which highlight the Xiaomi Mi Band 2 and VeryFit 2.0 devices as having acceptable validity, are therefore helpful. We bought the trackers as individual buyer on the consumer market. Researchers intending to use these in large-scale research cohorts may purchase these at an even lower cost in bulk. Another promising feature of the VeryFit 2.0 is that it has API (Application Programming Interface), which allows software developers to create custom software which integrates directly with the tracker (i.e. data from the tracker can be
sent automatically to the custom software). There is a growing trend for e-health and m-health research to use Fitbit and Garmin API (e.g. [71-73]). Therefore, validated low-cost tracker with API offer new data collection and intervention possibilities.

Our study included trackers with and without heart rate measurement. Trackers with highest validity all included heart rate measures, whereas those without showed lower validity. We can however not conclude from this study that the heart rate function increased validity. Studies testing the same model with and without heart rate function, and assessing the validity of the heart rate measurement in se would be needed to make this claim. The price of included trackers ranged from around €25 to around €50. The prices of the most accurate types, VeryFit 2.0 and Xiaomi Mi Band 2, are situated in the middle of this range (€30 to €40). This renders two models that are very attractive and accessible to the general public. Price may thus not be the determining factor in the validity of the trackers: more expensive within this range is not necessarily better. On the other hand, we cannot conclude that price plays no role and that trackers even less expensive than those included here (<€25), may also be valid. Indeed, a study on pedometers provided for free as gadgets with cereal boxes found that those were not valid [74].

Whereas validity evidence from this study, for low-cost devices measuring steps, MVPA and TST is not unequivocally good across the devices, user experience is also extremely important. A device which has high validity may not necessarily have a positive user-experience. Future research examining the user experience of low-cost trackers (e.g. focusing on issues such as functionality, reliability, ease of use, both of the device itself, and its accompanying app) will be valuable. Our preliminary experiences suggest that the user-experience of the low-cost trackers may be less positive than that for high-cost trackers (e.g. we tended to experience fewer technical issues with Fitbits than with the other devices in the current study). It can be assumed that the higher price of the high-cost trackers is partly determined by the investments made by the manufacturer to improve the user experience and to better develop the app supporting the tracker. Moreover, the low-cost activity trackers appear most valid for measuring steps. Pedometers that count steps are available at an even lower cost, but unlike activity trackers offer little additional functionality (e.g. feedback, information, social support) in an accompanying app, and are considered less usable by people than activity trackers [75-77]. Further work to explore these issues more rigorously and in greater depth is warranted.

**Strengths and limitations**

A strength of our study is that it is the first to scrutinise the validity of low-cost trackers addressing an important gap in the scientific literature to date. Methodological strengths of the study are the relatively large number of devices that were tested using the same methodology (allowing direct comparison of devices’ performance), that were tested multiple metrics (steps, MVPA and sleep), and that efforts were made to minimise bias, e.g. by randomising the order in which participants wore the devices. Limitations included that our sample was relatively
young and healthy. Based on previous literature, it seems likely that validity for measuring steps is likely to be somewhat lower in older and clinical populations (e.g. obese) [78]. As already noted, our reference devices were research-grade accelerometers with known validity limitations of their own. Therefore, they represent convergent validity rather than criterion validity, and there is a risk we may be under-estimating the low-cost trackers’ true validity. A further limitation is that this is a fast-moving field with new devices continually entering and exiting the market. In particular, since our study started, the Xiaomi MiBand 2 is replaced by its successor, the MiBand 3. Therefore, it would be beneficial that future research continuously investigates the validity of new low-cost tracker and other emerging devices. Furthermore, having an insight into the used algorithms and used cut-offs would be beneficial.

Conclusions
This study was the first to examine the validity of low-cost trackers. It found that validity was strongest for measurement of steps, there was some evidence of validity for measurement of sleep, whereas validity for measurement of MVPA time was weak. Validity ranged between devices, with the Xiaomi having the highest validity for measurement of steps, and the VeryFit performing relatively strongly across both sleep and steps domains. The tested low-cost trackers hold promise for cost-efficient measurement of movement behaviours. Further research investigating the user-experience of low-cost devices and their accompanying apps is needed before these devices can be confidently recommended.

Conflicts of Interest
None declared

Abbreviations
SES: socio economic status
PA: physical activity
MVPA: moderate to vigorous physical activity
TST: total sleep time
MAPE: mean absolute percentage error

References
1. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. Journal of applied physiology. 2005;99(3):1193-204.
2. Cappuccio FP, Cooper D, D’elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. European heart journal. 2011;32(12):1484-92.
3. Cappuccio FP, D’Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes care. 2010;33(2):414-20.
4. Guallar-Castillón P, Bayán-Bravo A, León-Muñoz LM, Balboa-Castillo T, López-García E, Gutierrez-Fisac JL, et al. The association of major patterns of physical activity, sedentary behavior and sleep with health-related quality of life: a cohort study. Preventive medicine. 2014;67:248-54.

5. Magee L, Hale L. Longitudinal associations between sleep duration and subsequent weight gain: a systematic review. Sleep medicine reviews. 2012;16(3):231-41.

6. Maher CA, Mire E, Harrington DM, Staiano AE, Katzmarzyk PT. The independent and combined associations of physical activity and sedentary behavior with obesity in adults: NHANES 2003 - 06. Obesity. 2013;21(12):E730-E7.

7. Cassidy S, Chau JY, Catt M, Bauman A, Trenell Ml. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233 110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ open. 2016;6(3):e010038.

8. Chastin SF, Palarea-Albaladejo J, Dontje ML, Skelton DA. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach. PloS one. 2015;10(10):e0139984.

9. Kok G, Gottlieb NH, Peters G-JY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health psychology review. 2016;10(3):297-312.

10. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychology. 2009;28(6):690.

11. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. International Journal of Behavioral Nutrition and Physical Activity. 2015;12(1):31.

12. DeSmet A, De Bourdeaudhuij I, Chastin S, Crombez G, Maddison R, Cardon G. Adults’ preferences for behavior change techniques and engagement features in a mobile application to promote 24-hour patterns of physical activity, sleep and sedentary behavior: under review.

13. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health psychology. 2008;27(3):379.

14. Hickey AM, Freedson PS. Utility of consumer physical activity trackers as an intervention tool in cardiovascular disease prevention and treatment. Progress in cardiovascular diseases. 2016;58(6):613-9.

15. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology (Cambridge, Mass). 2008;19(6):838-45.

16. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. Journal of medical Internet research. 2016;18(9).
17. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Medicine and science in sports and exercise. 2016;48(3):457.
18. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC sports science, medicine and rehabilitation. 2015;7(1):24.
19. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open. 2016;6(7):e011243. doi: 10.1136/bmjopen-2016-011243.
20. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The Effectiveness of a Web-Based Computer-Tailored Physical Activity Intervention Using Fitbit Activity Trackers: Randomized Trial. Journal of medical Internet research. 2018;20(12):e11321.
21. Olsen HM, Brown WJ, Kolbe-Alexander T, Burton NW. A Brief Self-Directed Intervention to Reduce Office Employees' Sedentary Behavior in a Flexible Workplace. J Occup Environ Med. 2018 Oct;60(10):954-9. PMID: 30001255. doi: 10.1097/JOM.0000000000001389.
22. Vanhaelewyn B, De Marez L. imec. Digimeter 2017. Measuring digital media trends in Flanders. Leuven: Leuven: imec; 2017.
23. Jia Y, Wang W, Wen D, Liang L, Gao L, Lei J. Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. PeerJ. 2018;6:e5350.
24. Beenackers MA, Kamphuis CB, Giskes K, Brug J, Kunst AE, Burdorf A, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. International journal of behavioral nutrition and physical activity. 2012;9(1):116.
25. Lallukka T, Sares-Jäske L, Kronholm E, Sääksjärvi K, Lundqvist A, Partonen T, et al. Sociodemographic and socioeconomic differences in sleep duration and insomnia-related symptoms in Finnish adults. BMC public Health. 2012;12(1):565.
26. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Elsevier; 2016.
27. Sorensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: A systematic review and integration of definitions and models. Bmc Public Health. 2012 Jan 25;12. PMID: WOS:000301526600001. doi: Artn 80 10.1186/1471-2458-12-80.
28. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. Journal of medical Internet research. 2017 Jan 24;19(1):e27. PMID: 28119275. doi: 10.2196/jmir.6709.
29. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. Journal of medical Internet research. 2012 Jan 27;14(1):e19. PMID: 22357448. doi: 10.2196/jmir.1619.
30. Porter CE, Donthu N. Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. J Bus Res. 2006 Sep;59(9):999-1007. PMID: WOS:000241430700006. doi: 10.1016/j.jbusres.2006.06.003.
31. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Medicine. 2016;13(2):e1001953.
32. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The Use of Wearable Activity Trackers Among Older Adults: Focus Group Study of Tracker Perceptions, Motivators, and Barriers in the Maintenance Stage of Behavior Change. JMIR mHealth and uHealth. 2019;7(4):e9832. PMID: 30950807. doi: 10.2196/mhealth.9832.
33. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity. 2015;12(1):159.
34. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. BMC research notes. 2014;7(1):952.
35. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Keadle SK, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. Journal of Physical Activity and Health. 2015;12(2):149-54.
36. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PloS one. 2016;11(9):e0161224.
37. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. Medicine & Science in Sports & Exercise. 2016;48(8):1619-28.
38. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer-and research-grade accelerometers: a comparative study in free-living conditions. JMIR mHealth and uHealth. 2016;4(3).
39. Ferguson T, Rowlands AV, Olds T, Maher C. 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. 2015;12(1):42.
40. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. BMC research notes. 2016;9(1):448.
41. Mantua J, Gravel N, Spencer R. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors. 2016;16(5):646.
42. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiology & behavior. 2016;158:143-9.
43. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. Health psychology open. 2016;3(2):2055102916679012.
44. Brooke SM, An H-S, Kang S-K, Noble JM, Berg KE, Lee J-M. Concurrent validity of wearable activity trackers under free-living conditions. Journal of strength and conditioning research. 2017;31(4):1097-106.

45. Brage S, Brage N, Franks PW, Ekelund U, Wong M-Y, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. Journal of applied physiology. 2004;96(1):343-51.

46. Van De Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography—a systematic review. Journal of sleep research. 2011;20(1pt2):183-200.

47. Colon A. The Best Fitness Trackers for 2019. PCmag. 2018;30-01-2019.

48. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Frontiers in physiology. 2017;8:725.

49. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR mHealth and uHealth. 2018;6(4).

50. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. Journal of medical Internet research. 2016;18(10):e264.

51. Rowlands ANNV, Stone MR, Eston RG. Influence of Speed and Step Frequency during Walking and Running on Motion Sensor Output. 2007:716-27. doi: 10.1249/mss.0b013e318031126c.

52. Masurier GUYCLE, Lee SM, Tudor-locke C. Motion Sensor Accuracy under Controlled and Free-Living Conditions ABSTRACT.905-10. doi: 10.1249/01.MSS.0000064996.63632.10.

53. Masurier GUYCLE, Tudor-locke C. Comparison of Pedometer and Accelerometer Accuracy under Controlled Conditions. (13):867-71. doi: 10.1249/01.MSS.0000064996.63632.10.

54. Freedson P, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Medicine and science in sports and exercise. 1998;30:777-81. doi: 10.1097/00005768-199805000-00021.

55. Santos-Lozano A, Santin-Medeiros F, Cardon G, Torres-Luque G, Bailon R, Bergmeir C, et al. Actigraph GT3X: Validation and Determination of Physical Activity Intensity Cut Points. International Journal of Sports Medicine. 2013 Nov;34(11):975-82. PMID: WOS:000326818100007. doi: 10.1055/s-0033-1337945.

56. Rowlands ANNV, Stone MR, Eston RG. Influence of Speed and Step Frequency during Walking and Running on Motion Sensor Output. 2007 (6):716-27. doi: 10.1249/mss.0b013e318031126c.

57. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998 May;30(5):777-81. PMID: 9588623.

58. Rich C, Geraci M, Griffiths L, Sera F, Dezateux C, Cortina-Borja M. Quality control methods in accelerometer data processing: defining minimum wear time.
59. Troiano RP, Berrigan D Fau - Dodd KW, Dodd Kw Fau - Masse LC, Masse Lc Fau - Tilert T, Tilert T Fau - McDowell M, McDowell M. Physical activity in the United States measured by accelerometer. (0195-9131 (Print)).

60. Shin M, Swan P, Chow CM. The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. Sleep Sci. 2015 Jan-Mar;8(1):9-15. PMID: 26483937. doi: 10.1016/j.slsci.2015.02.003.

61. Malavolti M, Pietrobelli A, Dugoni M, Poli M, Romagnoli E, De Cristofaro P, et al. A new device for measuring resting energy expenditure (REE) in healthy subjects. Nutr Metab Cardiovas. 2007 Jun;17(5):338-43. PMID: WOS:000248666900003. doi: 10.1016/j.numecd.2005.12.009.

62. Landis Jr Fau - Koch GG, Koch GG. The measurement of observer agreement for categorical data. (0006-341X (Print)).

63. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions. 2017;8(September). doi: 10.3389/fphys.2017.00725.

64. Kim Y, Park I, Kang M. Convergent validity of the international physical activity questionnaire (IPAQ): meta-analysis. Public Health Nutr. 2013 Mar;16(3):440-52. PMID: 22874087. doi: 10.1017/S1368980012002996.

65. Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003 Aug;35(8):1381-95. PMID: 12900694. doi: 10.1249/01.MSS.0000078924.61453.FB.

66. Maher C, Ryan J, Ambrosi C, Edney SJ Bph. Users’ experiences of wearable activity trackers: a cross-sectional study. 2017;17(1):880.

67. Banda JA, Haydel K F, Davila T, Desai M, Bryson S, Haskell WL, et al. Effects of varying epoch lengths, wear time algorithms, and activity cut-points on estimates of child sedentary behavior and physical activity from accelerometer data. 2016;11(3):e0150534.

68. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille SJ M, et al. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. 2013;45(5):964.

69. Brickwood KJ, Watson G, O’Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. Jmir Mhealth and Uhealth. 2019 Apr 12;7(4). PMID: WOS:000465347200001. doi: ARTN e11819 10.2196/11819.

70. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for Sustained Use of an Activity Tracker: Observational Study. Jmir Mhealth and Uhealth. 2017 Oct;5(10). PMID: WOS:000415044800004. doi: ARTN e164 PMID 29084709 10.2196/mhealth.7311.
71. Shaw R, Levine E, Streicher M, Strawbridge E, Gierisch J, Pendergast J, et al. Log2Lose: development and lessons learned from a mobile technology weight loss intervention. 2019;7(2):e11972.
72. Krans M, van de Wiele L, Bullen N, Diamond M, van Dantzig S, de Ruyter B, et al., editors. A group intervention to improve physical activity at the workplace. International Conference on Persuasive Technology; 2019: Springer.
73. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: The ACTIVATE Trial. 2019.
74. De Cocker K, Cardon G Fau - De Bourdeaudhuij I, De Bourdeaudhuij I. Validity of the inexpensive Stepping Meter in counting steps in free living conditions: a pilot study. (1473-0480 (Electronic)).
75. Husted HM, Llewellyn TL. The Accuracy of Pedometers in Measuring Walking Steps on a Treadmill in College Students. International journal of exercise science. 2017;10(1):146-53. PMID: 28479955.
76. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR mHealth and uHealth. 2016;4(1):e7-e. PMID: 26818775. doi: 10.2196/mhealth.4225.
77. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. J Med Internet Res. 2018 Mar 22;20(3):e110. PMID: 29567635. doi: 10.2196/jmir.9157.
78. Kenyon A, McEvoy M, Sprod J, Maher CJ AoPM, Rehabilitation. Validity of pedometers in people with physical disabilities: A systematic review. 2013;94(6):1161-70.