A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals

Middelweerd, A.; van der Ploeg, H.P.; van Halteren, A.; Twisk, J.W.R.; Brug, J.; te Velde, S.J.

DOI
10.1249/MSS.0000000000001225

Publication date
2017

Document Version
Final published version

Published in
Medicine and science in sports and exercise

Citation for published version (APA):
Middelweerd, A., van der Ploeg, H. P., van Halteren, A., Twisk, J. W. R., Brug, J., & te Velde, S. J. (2017). A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals. *Medicine and science in sports and exercise*, *49*(6), 1270-1279. https://doi.org/10.1249/MSS.0000000000001225

OPEN

A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals

ANOUK MIDDELWEERD[1], HIDDE P. VAN DER PLOEG[2], AART VAN HALTEREN[3,4], JOS W. R. TWISK[1], JOHANNES BRUG[5], and SASKIA J. TE VELDE[1]

[1]Department of Epidemiology and Biostatistics and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, THE NETHERLANDS; [2]Department of Public and Occupational Health and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, THE NETHERLANDS; [3]Department of Computer Science, VU University Amsterdam, THE NETHERLANDS; [4]Philips Research, Eindhoven, THE NETHERLANDS; and [5]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, THE NETHERLANDS

ABSTRACT

MIDDELWEERD, A., H. P. VAN DER PLOEG, A. VAN HALTEREN, J. W. R. TWISK, J. BRUG, and S. J. TE VELDE. A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals. *Med. Sci. Sports Exerc.*, Vol. 49, No. 6, pp. 1270–1279, 2017. **Purpose**: Accelerometer-based wearables can provide the user with real-time feedback through the device's interface and the mobile platforms. Few studies have focused on the minute-by-minute validity of wearables, which is essential for high-quality real-time feedback. This study aims to assess the validity of the Fitbit One compared with the ActiGraph GT3x + for assessing physical activity (i.e., steps, time spent in moderate, vigorous, and moderate–vigorous physical activity) in young adults using traditional time intervals (i.e., days) and smaller time intervals (i.e., minutes and hours). **Methods**: Healthy young adults (N = 34) wore the ActiGraph GT3x+ and a Fitbit One for 1 wk. Three aggregation levels were used: minute, hour, and day. Mixed models analyses, intraclass correlation coefficients, Bland–Altman analyses, and absolute error percentage for steps per day were conducted to analyze the validity for steps and minutes spent in moderate, vigorous, and moderate–vigorous physical activity. **Results**: As compared with ActiGraph (mean = 9 steps per minute, 509 steps per hour and 7636 steps per day), the Fitbit One systematically overestimated physical activity for all aggregation levels: on average 0.82 steps per minute, 45 steps per hour, and 677 steps per day. Strong and significant associations were found between ActiGraph and Fitbit results for steps taken (B = 0.72–0.89). Weaker but statistically significant associations were found for minutes spent in moderate, vigorous, and moderate–vigorous physical activity for all time intervals (B = 0.39–0.57). **Conclusions**: Although the Fitbit One overestimates the step activity compared with the ActiGraph, it can be considered a valid device to assess step activity, including for real-time minute-by-minute self-monitoring. However, agreement and correlation between ActiGraph and Fitbit One regarding time spent in moderate, vigorous, and moderate–vigorous physical activity were lower. **Key Words:** FITBIT ONE, VALIDATION, MINUTE-BY-MINUTE, DAILY LIFE, REAL-TIME FEEDBACK, SELF-MONITORING

Commercial wearable technologies that continuous monitor physical activity may be helpful tools for (self-) monitoring and for providing feedback on physical activity behaviors. For example, the Fitbit One is an accelerometer-based wearable activity tracking device that is easy to use and has a user-friendly interface. The activity monitor wirelessly uploads the activity data to the user's account, which is accessed through a smartphone application or website (11,18). It assesses daily activities, allows the user to monitor his/her progress, and provides real-time feedback (e.g., current amount of steps taken, stairs taken and a growing or shrinking flower depicts the activity level) to encourage the user to be more physically active (11). In addition, the Fitbit One and its web interface include behavior change techniques typically used in physical activity interventions (17). The Fitbit One—like other similar wearables—may offer new opportunities in interventions that target physical activity promotion.

Several studies examined the validity of different types of Fitbit activity trackers (7,20,22,24); however, as far as we are aware, only eight studies examined the validity of the Fitbit One. So far, five studies used a controlled laboratory setting to examine the validity of the Fitbit One with respect to steps taken (n = 4) and energy expenditure (n = 1) (3,16,21,22). Overall, in laboratory settings, studies reported acceptable

Address for correspondence: Hidde P. Van der Ploeg, Ph.D., Department of Epidemiology and Biostatistics, VU Medical Center, Amsterdam De Boelelaan 1089a, Amsterdam 1081HV, The Netherlands; E-mail: hp.vanderploeg@vumc.nl.
Submitted for publication April 2016.
Accepted for publication January 2017.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/17/4906-1270/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000001225

percentages of relative errors (1.3%–10.5%) for number of steps taken with speeds of 0.7–1.78 $m{\cdot}s^{-1}$ on a treadmill (21–23). Moreover, the Fitbit underestimated the energy expenditure compared with indirect calorimetry. So far, only three studies examined the validity of the Fitbit One in daily life (10,14,20). High Pearson's r and intraclass correlations were reported for steps taken per day compared with the ActiGraph (10). However, the Fitbit One was less valid in assessing daily minutes spent in moderate–vigorous physical activity (10,20). These findings suggest that the Fitbit One is a valid instrument to assess daily step activity but less valid to assess daily minutes spent in moderate–vigorous physical activity.

Moreover, the Fitbit One can be used as a stand-alone intervention or as an intervention tool to monitor daily life physical activity and to provide real-time feedback. It can be used with its own applications or in combination with a third-party smartphone application. The Fitbit website provides access to minute-by-minute activity data through the application programming interface (API). A recent study in overweight, postmenopausal women showed promising results in favor of the intervention group that received the Fitbit One as an intervention tool (2). Because real-time feedback is and should be based on data collected throughout the day and thus on data collected over smaller time intervals than the whole day, it is important to know the validity of the Fitbit for levels of physical activity based on smaller time intervals than the whole day (i.e., total daily activity). This is important because even if studies have indicated that the Fitbit One is valid in assessing daily activity, this could be because measurement errors—over- and underestimations—can balance out throughout the day which would result in good daily validity but not necessarily good minute or hourly validity. Therefore, it is important to know the validity of the Fitbit using smaller time intervals (i.e., per minute and hour) and, thus, if the Fitbit One can be used for real-time feedback. This focus on the smaller time intervals is important as a poor hourly or minute-by-minute validity might be off-putting to users that experience the wearable is not matching their activities in real time, and consequently, it might be less suitable for physical activity interventions.

One previous study looked at 3-h periods and demonstrated a good concurrent validity of the Fitbit Ultra—the forerunner of the Fitbit One—to assess physical activity in daily life in patients with chronic obstructive pulmonary disease (27). Another study used smaller time intervals, i.e., minutes, hours, and days, to assess the interdevice reliability, but the study only had one participant (7). Therefore, to date, little is known about the validity of the Fitbit One in detecting physical activity using smaller time intervals relevant for real-time feedback and instant behavioral insights to its users. Hence, the primary aim of the current study is to assess the construct validity of the Fitbit One for steps and minutes spent in moderate physical activity, vigorous physical activity, and moderate–vigorous physical activity for minute and hourly intervals against the ActiGraph. The secondary aim was to assess the more traditional construct validity of the Fitbit One for steps, minutes spent in moderate physical activity, vigorous physical activity, and moderate–vigorous physical activity per day against the ActiGraph.

METHODS

Participants. A sample of 34 participants (23 females and 11 males) agreed to participate in the current study. Participants were actively recruited in Amsterdam by flyers distributed on one of the main university campuses, and by e-mail, and direct person-to-person communication at this university. Participants were eligible when they met the following criteria: (a) age 19–30 yr, (b) owned a smartphone, (c) fluent in Dutch or English, and (d) signed the informed consent form.

Procedure. Participants were asked to visit the study center to pick up the instructions and devices. After providing informed consent, the participants completed an online questionnaire collecting demographic information such as age, gender, self-reported height, and weight. The participants received log-in information for the Fitbit website and instructions on how to use the devices; for example, how to synchronize the Fitbit, and how to wear the devices. Participants were instructed to wear both activity monitors on the right hip using an elastic belt for seven consecutive days during waking hours. Furthermore, they were instructed to remove the activity monitors during water activities and sleeping. After the assessment week, participants returned the activity monitors and received a 10-euro voucher as incentive for participating.

The study protocol was approved by the Medical Ethical Committee of the VU University Medical Centre Amsterdam.

Measurements. The ActiGraph GT3X+ (ActiGraph Inc., Pensacola, FL) was used as the reference method (construct validity) to compare the Fitbit with. The ActiGraph is recognized as a reasonable valid tool to assess physical activity objectively in adults and has been used in numerous studies (1,28). The ActiGraph GT3X+ is a triaxial accelerometer that is able to convert accelerations to step counts. The monitors were set to collect raw data at 100 Hz. After data were aggregated into 1-min intervals, Troiano's definitions—using the vertical counts—were used to calculate time spent in sedentary (<100 counts per minute), light (100–2019 counts per minute), moderate (2020–5998 counts per minute), vigorous ($5999 \leq$ counts per minute), and moderate to vigorous (moderate–vigorous physical activity, $2020 \leq$ counts per minute) physical activities (24).

The Fitbit One (Fitbit Inc., San Francisco, CA) is a lightweight triaxial accelerometer with a build in altitude monitor. The Fitbit collects data in 1-min intervals. Data were retrieved using the open API and saved as xml file. Fitbit uses an algorithm to detect motion patterns and to convert these accelerations to step counts (12). Fitbit calculates the intensity of the activity by means of an algorithm only known within the Fitbit company. On the basis of the intensity of the activity, Fitbit then estimates the corresponding MET values to classify the intensity in sedentary, lightly active, fairly active, and very active (13). The monitor can be worn in the front pocket, on a belt or a bra; however, for the present study, participants were asked to wear the Fitbit on the right hip with a waist belt, next to the ActiGraph.

Participants were asked to record their wear time through a daily e-mail with a link to a questionnaire recalling the previous day. This e-mail was sent at the beginning of each day—starting at the second day.

Data handling. First, ActiGraph data were checked for nonwear time using Actilife 6.0 (ActiGraph Inc.). Troiano's definitions were used to identify nonwear time; for example, periods with consecutive strings of zero's for at least 60 min (24). In addition, up to two interruption intervals of <100 counts that appeared in the middle of long strings of zero-count intervals were filtered out (5,9). Intervals with >20,000 counts per minute were considered to be spurious (5,19).

Second, Fitbit data were matched to the periods of wear time of the ActiGraph data. All Fitbit data with zero steps per hour were compared with the participant's reports and if necessary excluded from the analyses. Some participants reported low battery of the Fitbit, and those time intervals were deleted. For both monitors, at least 4 d of minimal 10 h of wear time was required to for a participant to be included in the analyses.

Data analyses. In preparation for the analyses, three levels of aggregation were distinguished (i.e., minutes, hours, and days). For each aggregation level and each type or intensity of activity (i.e., steps, minutes spent in moderate, vigorous, and moderate–vigorous activities), the validity was examined in five steps.

First, the absolute error percentage [100(|Fitbit − ActiGraph|)/ActiGraph] was calculated to provide an indicator of the overall difference between the Fitbit and the ActiGraph. Because the absolute error percentage cannot be calculated for ActiGraph values of zero or close to zero, the error percentage was calculated for steps per day only.

Second, because longitudinal data observations within one subject over time are correlated, systematic differences were obtained by means of linear mixed model analyses with a four-level structure (minute measurements were clustered within hours, within days, and within weeks). The dependent variable was the continuous activity data assessed by either the Fitbit or the ActiGraph, and the independent dichotomous variable was the device (ActiGraph = 0, Fitbit = 1). The obtained regression coefficient represents the systematic difference between the Fitbit and the ActiGraph adjusted for the nested design.

Third, Bland–Altman plots, including the limits of agreement (LoA), were used to provide a visual representation of the systematic differences and to assess potential nonsystematic differences between the ActiGraph and the Fitbit. As the activity data in the current study comprise a mixture of between and within individual information on the differences between the two measurements, we applied the method proposed by Bland and Altman to estimate the LoA for repeated measures by adjusting for the nonindependence of the observations. Because of the fact that Fitbit and ActiGraph collect data in 1-min epochs, the variable for time spent in moderate, vigorous, or moderate–vigorous physical activity at the minute level was dichotomous, either that minute was at that intensity or not. Therefore, Bland–Altman analyses were not conducted for minutes spent in moderate, vigorous, and moderate–vigorous physical activity for the aggregation level minute.

Fourth, intraclass correlation coefficients (ICC) for continuous data were used as parameters for criterion validity of the Fitbit compared with the ActiGraph. The ICC values (two-way random effects model with an absolute agreement definition) were calculated for all aggregation levels for step activity and time spent per intensity level. For the present study, an ICC ≥0.75 was defined as "excellent," an ICC between 0.4 and 0.75 was defined as "fair to good" and an ICC <0.4 as "poor" (4,5).

Lastly, the association between ActiGraph data (dependent variable) and Fitbit data (independent variable) was also obtained with linear mixed model analyses with the same four-level structure as described previously. Regression coefficients were interpreted as "excellent" (B = 0.8–1.0), "substantial" (B = 0.6–0.8), "moderate" (B = 0.4–0.6), "fair" (B = 0.2–0.4), and as "poor" (B = 0.0–0.2) correlations (8).

All analyses were checked for outliers (≥3 SD of the residuals) and when necessary sensitivity analyses was conducted without outliers. Results of the sensitivity analyses can be found in Appendix 1 (see Table, Supplemental Digital Content 1, Results of the analyses that were used to assess the agreement, association and correlation of the Fitbit compared with the ActiGraph for the four types of activity data, http://links.lww.com/MSS/A865). Statistical analyses were performed in R Studio (R Studio, Boston) using lme, lmer, glm, and ggplot2 packages.

RESULTS

Descriptive characteristics. Aggregating the activity data into the three aggregation levels resulted in data sets consisting of 138,224 observations for data aggregated into minute intervals, 2518 observations for data aggregated into hour intervals, and 168 observations for data aggregated into day intervals.

Table 1 shows descriptive characteristics of the study sample (N = 34); the majority was female and of normal weight. Three participants were excluded from the analyses because of the lack of eligible data (at least 4 d of 10 h), and one participant did not return the measurement devices. The remaining 30 participants wore the devices for 13.7 (SD = 1.9) $h \cdot d^{-1}$ on average.

For steps per day, the absolute mean error percentages was 11.4% (median = 7.7%). As mentioned previously, this analysis was only applicable for steps per day because of ActiGraph values of zero or close to zero for the other measurements.

Mixed model analyses showed a systematic overestimation of the Fitbit for all activity data. The analyses for aggregation level minute showed a significant overestimation of 0.82

TABLE 1. Descriptive information (mean ± SD) for the participants and those included and excluded from the analyses.

	All Participants	Included from Analyses	Excluded from Analyses[a]
N	34	30	4
Gender (*n* = females)	23	20	3
Age (yr)	23.9 ± 3.9	23.9 ± 3.9	23.8 ± 5.2
Self-reported height (cm)	174.8 ± 9.9	173.8 ± 9.3	182.5 ± 12.0
Self-reported weight (kg)	66.5 ± 8.9	65.2 ± 7.5	76.3 ± 14.1
Self-reported body mass index ($kg \cdot m^{-2}$)	21.7 ± 2.2	21.6 ± 2.2	22.7 ± 2.0

[a]Excluded from analyses are participants with ineligible data (*n* = 3) or who dropped out (*n* = 1).

(95% confidence interval [CI] = 0.64–1.00, mean ActiGraph = 9.28) steps per minute (Table 2). The analyses for the hour level aggregation showed a significant overestimation of 45.16 (95% CI = 3.12–87.20, mean ActiGraph = 509.46) steps, 2.86 (95% CI = 2.55–3.16, mean ActiGraph = 2.30) minutes of moderate physical activity, 1.03 (95% CI = 0.86–1.19, ActiGraph = 0.29) minutes of vigorous physical activity, and 3.88 (95% CI = 3.49–4.27, ActiGraph = 2.59) minutes of moderate–vigorous physical activity per hour. The analyses for day level aggregation showed a significant overestimation of 42.82 (95% CI = 36.29–49.34, ActiGraph = 34.49) minutes of moderate physical activity, 15.36 (95% CI = 12.29–18.44, ActiGraph = 4.29) minutes of vigorous physical activity, and 58.18 (95% CI = 50.11–66.25, ActiGraph = 38.78) minutes of moderate–vigorous physical activity per day.

The Bland–Altman plots provide a visual representation of the systematic differences and the plots are shown in Figures 1–3. All Bland–Altman plots (Figs. 1–3) show wide ranges for the LoA. For the assessment of steps per day, per hour, and per minute, the differences between the two measurements (*y*-axis) seem to be similar for different mean values of the two measurements (*x*-axis) (Fig. 1). However, a smaller range in differences in steps is seen for ≥100 mean steps per minute. By contrast, the physical activity intensities tend to show a larger range in differences between the two methods (*y*-axis) with larger mean values, especially for minutes in moderate–vigorous activity per day (Fig. 3C), and in minutes in moderate activity per hour (Fig. 2A) and per day (Fig. 3A). Finally, there seems to be a positive bias for the comparison of the two methods regarding time spent in moderate–vigorous physical activity per hour (Fig. 2C) as well as per day (Fig. 3C). This is reflected by the more positive different scores (*y*-axis) for the larger mean values (*x*-axis). Table 2 shows the results of all Bland–Altman analyses, mixed model analyses, and ICCs.

Analyses for the minute level aggregation showed an excellent ICC for steps per minute (ICC = 0.80). Analyses for the hour level aggregation showed an excellent ICC for steps per hour (ICC = 0.97), but ICC values for time spent on moderate, vigorous activity, and moderate–vigorous physical activity were much lower (ICC = 0.57–0.67). Analyses for the day level aggregation showed an excellent ICC for steps per day (ICC = 0.96), but ICC values for time spent on moderate, vigorous activity, and moderate–vigorous physical activity were much lower (ICC = 0.33–0.46).

TABLE 2. Mean values of the ActiGraph data and results of the analyses that were used to assess the agreement, association, and correlation of the Fitbit compared with the ActiGraph for the four types of activity data (e.g., steps, moderate, vigorous, and moderate–vigorous activity) and four time intervals (i.e., minute, *n* = 138,224; hour, *n* = 2518; day, *n* = 168).

Outcome Measurement		Agreement Analyses			Association and Correlation Analyses		
Type of Activity Data	Time Interval	ActiGraph (Mean ± SD)	Fitbit (Mean ± SD)	Bland–Altman Analyses, Mean Difference (LoA)	Systematic Differences with Mixed Models Analyses, *B* (95% CI)	Association with Mixed Models Analyses, *B* (95% CI)	Intraclass Correlation (95% CI)[a]
Steps	Minute	9.28 ± 23.42	10.10 ± 25.97	0.82 (−30.72 to 32.36)	**0.82(0.64 to 1.00)[b]**	**0.72 (0.72 to 0.72)[b]**	0.80 (0.79 to 0.80)[a]
	Hour	509.46 ± 744.77	554.62 ± 814.99	45.16 (−320.89 to 411.21)	*45.16 (3.12 to 87.20)[c]*	**0.89 (0.89 to 0.90)[c]**	0.97 (0.97 to 0.98)[a]
	Day	7635.79 ± 4553.39	8312.65 ± 5135.68	676.86 (−1584.09 to 2937.82)	676.86 (−152.78 to 1506.50)[d]	**0.86 (0.84 to 0.89)[d]**	0.96 (0.91 to 0.98)[a]
Moderate activity	Minute	0.04 ± 0.20	0.09 ± 0.29				
	Hour	2.30 ± 4.86	5.16 ± 6.33	2.86 (−6.73 to 12.44)	**2.86 (2.55 to 3.16)[c]**	**0.51 (0.49 to 0.54)[c]**	0.57 (0.33 to 0.71)[a]
	Day	34.49 ± 26.94	77.30 ± 42.79	42.82 (−24.38 to 110.01)	**42.82 (36.29 to 49.34)[d]**	**0.39 (0.31 to 0.47)[d]**	0.33 (−0.09 to 0.61)[a]
Vigorous activity	Minute	0.01 ± 0.07	0.02 ± 0.15				
	Hour	0.29 ± 2.40	1.31 ± 3.66	1.03 (−3.82 to 5.87)	**1.03 (0.86 to 1.19)[a]**	**0.50 (0.48 to 0.51)[c]**	0.66 (0.54 to 0.74)[a]
	Day	4.29 ± 11.55	19.65 ± 20.00	15.36 (−11.66 to 42.38)	**15.36 (12.29 to 18.44)[d]**	**0.44 (0.38 to 0.50)[d]**	0.46 (−0.05 to 0.72)[a]
Moderate–vigorous activity	Minute	0.05 ± 0.21	0.12 ± 0.32				
	Hour	2.59 ± 5.72	6.47 ± 8.42	3.88 (−5.88 to 13.64)	**3.88 (3.49 to 4.27)[c]**	**0.57 (0.56 to 0.59)[c]**	0.67 (0.28 to 0.83)[a]
	Day	38.78 ± 31.96	96.96 ± 54.68	58.18 (−12.66 to 129.02)	**58.18 (50.11 to 66.25)[d]**	**0.46 (0.41 to 0.52)[d]**	0.37 (−0.10 to 0.69)[a]

n, the number of observation for the aggregated data sets per time interval; *B*, regression coefficient; 95% CI, 95% confidence interval (B ± 1.96 SE). Bland–Altman analyses show the mean difference between the Fitbit One and the ActiGraph (Fitbit One–ActiGraph) and LoA (mean difference ± 2 SD). Bold text indicates $P = 0.000$. Italic text indicates $P = 0.01$.

[a]ICC values denote the interdevice reliability of the Fitbit and ActiGraph and was used for continuous variables.

[b]Adjustments for clustering within individuals, days and hours for the mixed models analyses.

[c]Adjustments for clustering within individuals and days for the mixed models analyses.

[d]Adjustments for clustering within individuals for the mixed models analyses.

Mixed model analysis for the aggregation level minute were conducted for steps only and showed a substantial association (B = 0.72, 95% CI = 0.79–0.80). Analyses for aggregation level hour showed an excellent association for steps (B = 0.89, 95% CI = 0.89–0.90) but moderate associations for time spent per intensity level (B = 0.51–0.57). Analyses for the aggregation level day showed an excellent association for steps (B = 0.86, 95% CI = 0.84–0.89), moderate association for time spent in vigorous and moderate–vigorous physical activity (B = 0.46–0.51), and a poor association for time spent in moderate physical activity (B = 0.38, 95% CI = 0.31–0.46).

The sensitivity analyses without outliers showed smaller systematic errors for steps per day, hour, and minute compared with analyses with outliers.

DISCUSSION

This study compared Fitbit One measurements with ActiGraph GT3X+ in terms of assessing physical activity in different time intervals to assess the construct validity of the Fitbit One for providing real-time physical activity feedback. The results indicate that the Fitbit One systematically overestimates physical activity, shows excellent associations with ActiGraph measures for step counts, but only fair to good associations for time spent in moderate, vigorous physical activity, and combined moderate–vigorous physical activity. These measurement properties of the Fitbit One were similar in the small time intervals (i.e., minutes and hours) and the more aggregated time intervals (i.e., days), which suggests that the Fitbit One is well suited for providing real-time feedback on steps activity for self-monitoring.

The primary aim of this study was to assess the construct validity of the Fitbit One using smaller time intervals compared with earlier studies. This focus on the smaller time intervals is important because even if the Fitbit One was shown to be valid in assessing daily activity, this could be because measurement errors, over- and underestimations, can balance out throughout the day, which would results in good daily validity but not necessarily good minute or hourly validity. Wearables, such as the Fitbit One, are often used to provide the user with real-time feedback for self-monitoring purposes, which is increasingly used in physical activity interventions. Hence, minute-by-minute validity is important as poor validity might be off-putting to users that experience the wearable is not matching their activities in real time. Therefore, it is important to also know the validity of the Fitbit using smaller time intervals and, thus, if the Fitbit One can be used for real-time feedback throughout the day. The results showed a substantial correlation and significant association for steps per minute and an excellent correlation and significant association for steps per hour of the Fitbit One and ActiGraph. Thus, results for the assessment of step activity were similar for the aggregation levels hour and day, whereas the results for steps per minute were slightly slower but sill good. However, for minutes spent in moderate, vigorous, and moderate–vigorous physical activity per hour, the Fitbit showed moderate correlations. Furthermore, the Fitbit systematically overestimates activity compared with the ActiGraph in all analyses. Looking at the systematic differences, there was no evidence for influence of time because the overestimation is approximately the same for all time intervals. Consequently, we conclude that the Fitbit is suitable for providing real-time feedback on progress in step activity and can thus be used as an intervention tool to monitor step activity and to provide feedback throughout the day. From a validity perspective, this makes the Fitbit One suitable as a self-monitoring tool in physical activity interventions, especially when using steps. However, the Fitbit is less suitable for providing instant real-time feedback for physical activity intensity levels.

The secondary aim of this study was to assess the construct validity of the Fitbit One for daily physical activity. Similar results were found as for the analyses with smaller time intervals. An excellent correlation and significant associations were seen for daily step activity; however, time per intensity level varied substantially among the Fitbit and ActiGraph. These results are in line with Ferguson et al. (10) who reported a high Pearson correlation for step activity (r = 0.99) and minutes of moderate–vigorous activity (r = 0.91) and an excellent interdevice agreement for step activity (ICC = 0.95) of the Fitbit One and ActiGraph, but a lower value for minutes of moderate–vigorous activity (ICC = 0.46). The Fitbit substantially overestimates daily step activity and daily minutes spent in moderate, vigorous, and moderate–vigorous physical activity activities compared with the ActiGraph, which is in line with previous research (8,10). However, Gomersall et al. (14) reported that the Fitbit One underestimated daily minutes of daily minutes of moderate–vigorous physical activity with 19.2 min·day^{-1}. On the basis of these findings—when being compared with the ActiGraph, the Fitbit is not suitable to assess intensity levels of physical activity in daily life. The substantial differences between the two measurements could have arisen from differences in the algorithms that converted the accelerometers activity counts into intensity levels of physical activity. Unfortunately, Fitbit's algorithm is not publicly available, and therefore it remains unclear how the Fitbit calculates the intensity levels of physical activity. However, it should be noted that in a validation study of Lee et al. (16)—in a laboratory setting—lower mean absolute error percentages were found between energy expenditure assessed by indirect calorimetry and estimated energy expenditure by the Fitbit One than energy expenditure estimated by the ActiGraph. These results indicate, that the Fitbit One might be more suitable for assessing energy expenditure than the ActiGraph in a laboratory setting.

Although the Fitbit systematically overestimates step activity compared with the ActiGraph, according to Tudor-Locke et al. (25), the median of the absolute error percentage is within the range of acceptable error. In general, an error percentage of 10% is considered to be acceptable for assessing step activity. Thus, the acceptable systematic differences between and the high correlations of the Fitbit and the

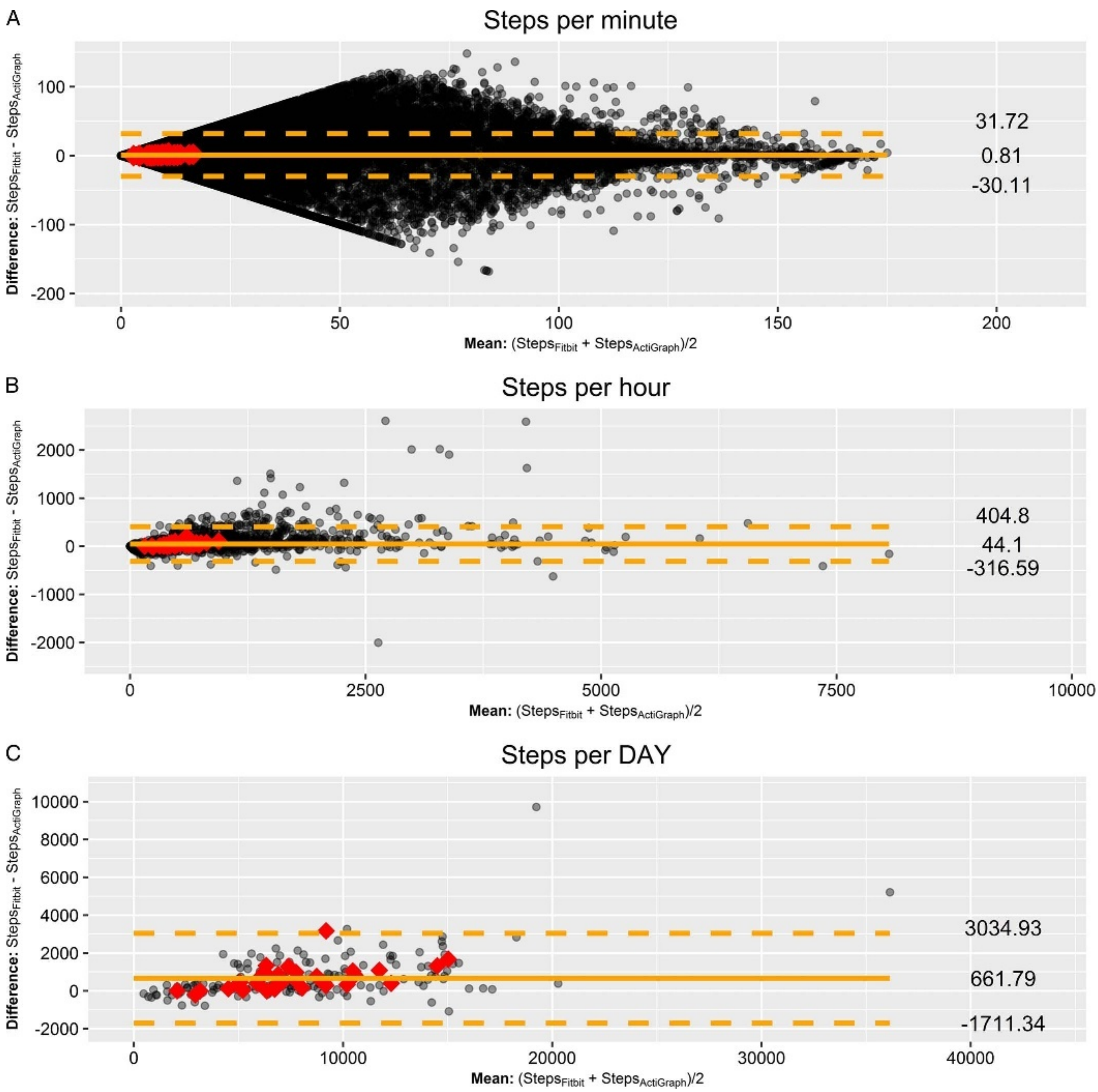

FIGURE 1—Bland–Altman plot for steps per day, hour, and minute for measurements with the ActiGraph and Fitbit, with adjusted means and LoA. A, Bland–Altman plot for steps per minute. B, Bland–Altman plot for steps per hour. C, Bland–Altman plot for steps per day. The *x*-axis shows the average number of steps, and the *y*-axis shows the differences in steps between the two methods. The mean values of all steps per day, hour, and minute per individual are plotted in *red*.

ActiGraph strengthen the hypothesis that the Fitbit is suitable for assessing daily step activity and that it could replace more expensive accelerometers in physical activity interventions. Nevertheless, an overestimation of 677 steps per day by the Fitbit One should be taken into account. Furthermore, it should be noted that although accelerometers are commonly used for the assessment of physical activity in daily life, they have some limitations. Accelerometers are limited in their ability to capture upper body movement, and cycling (28). Further, accelerometers are not waterproof and thus they are not able to record water activities (28). Furthermore, accelerometers also underestimate the energy expenditure for activities while carrying heavy loads because the acceleration patterns remain unchanged, despite the loads carried. Thus, accelerometers such as the ActiGraph and Fitbit One might underestimate total daily physical activity and energy expended (28).

A strength of this study is the novel approach by assessing the validity of the Fitbit for different time intervals. Although

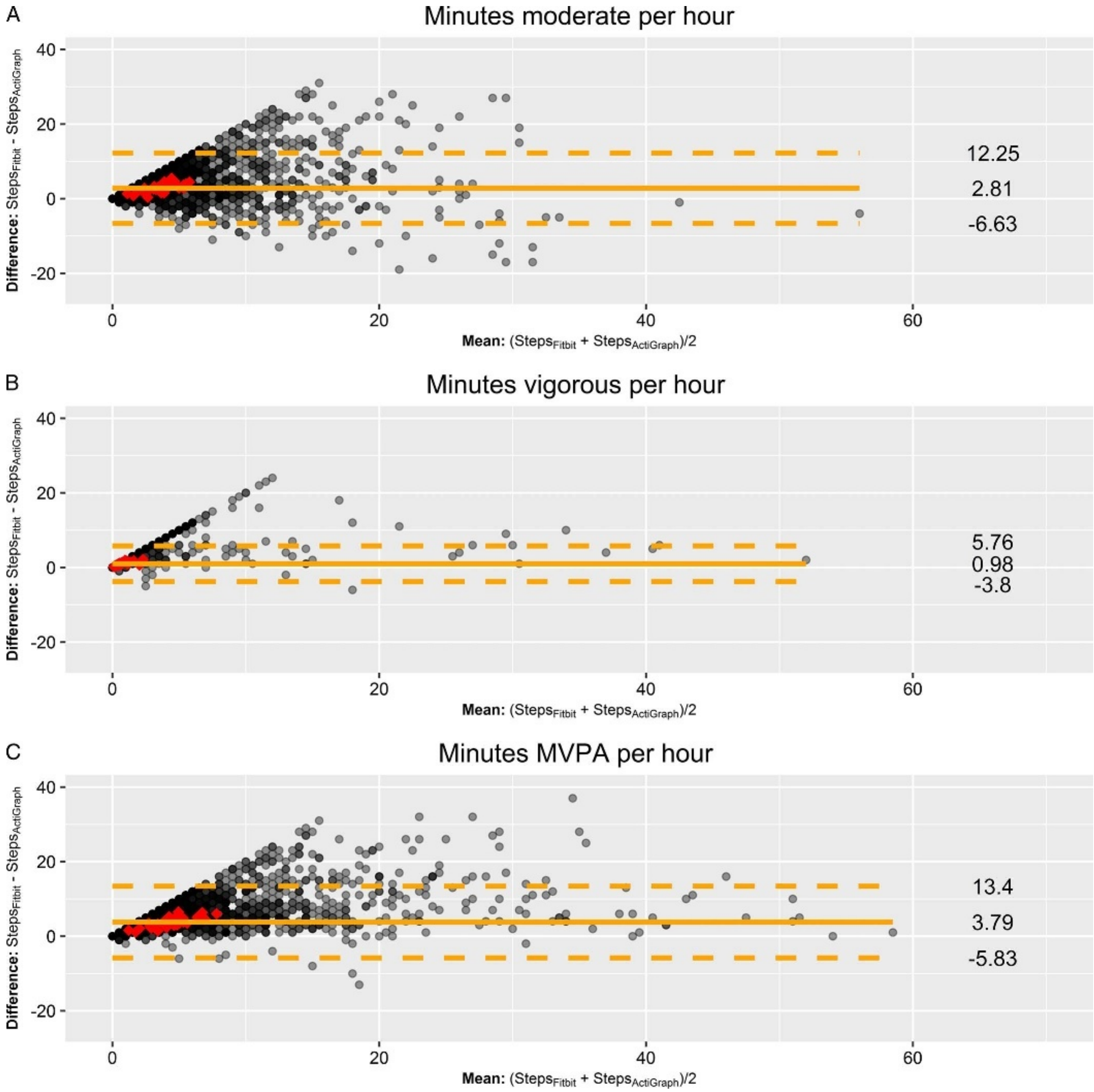

FIGURE 2—Bland–Altman plot for minutes of moderate physical activity, vigorous physical activity, and moderate to vigorous physical activity (MVPA) per hour, for measurements with the ActiGraph and Fitbit, with adjusted means and LoA. A, Bland–Altman plot for minutes of moderate physical activity per hour. B, Bland–Altman plot for minutes of vigorous physical activity per hour. C, Bland–Altman plot for minutes of moderate to vigorous physical activity per hour. The *x*-axis shows the average of the two measurements, and the *y*-axis shows the differences between the two methods. The mean values of all steps per day, hour, and minute per individual are plotted in *red*.

several studies have assessed the validity of the Fitbit One, our study did not only assess the validity of the Fitbit for total daily physical activity but focused on smaller time intervals to determine the validity of the Fitbit for providing real-time feedback for self-monitoring as well. Furthermore, this study used linear mixed models analyses—and thus adjusted for clustering—to determine the correlations and systematic differences for the measurements of the Fitbit and ActiGraph in addition to the traditional analyses (e.g., Bland–Altman analyses and ICC), which do not adjust for the dependency of the data. In addition, the protocol was designed to be comparable with previous validation studies (6,15,26). Lastly, this study used the API to obtain the Fitbit data, instead of manually copying the displayed data on the website or app that is prone to errors. This study also has some limitations. First, the ActiGraph GT3x+ was used as a reference measurement, and the limitations of this

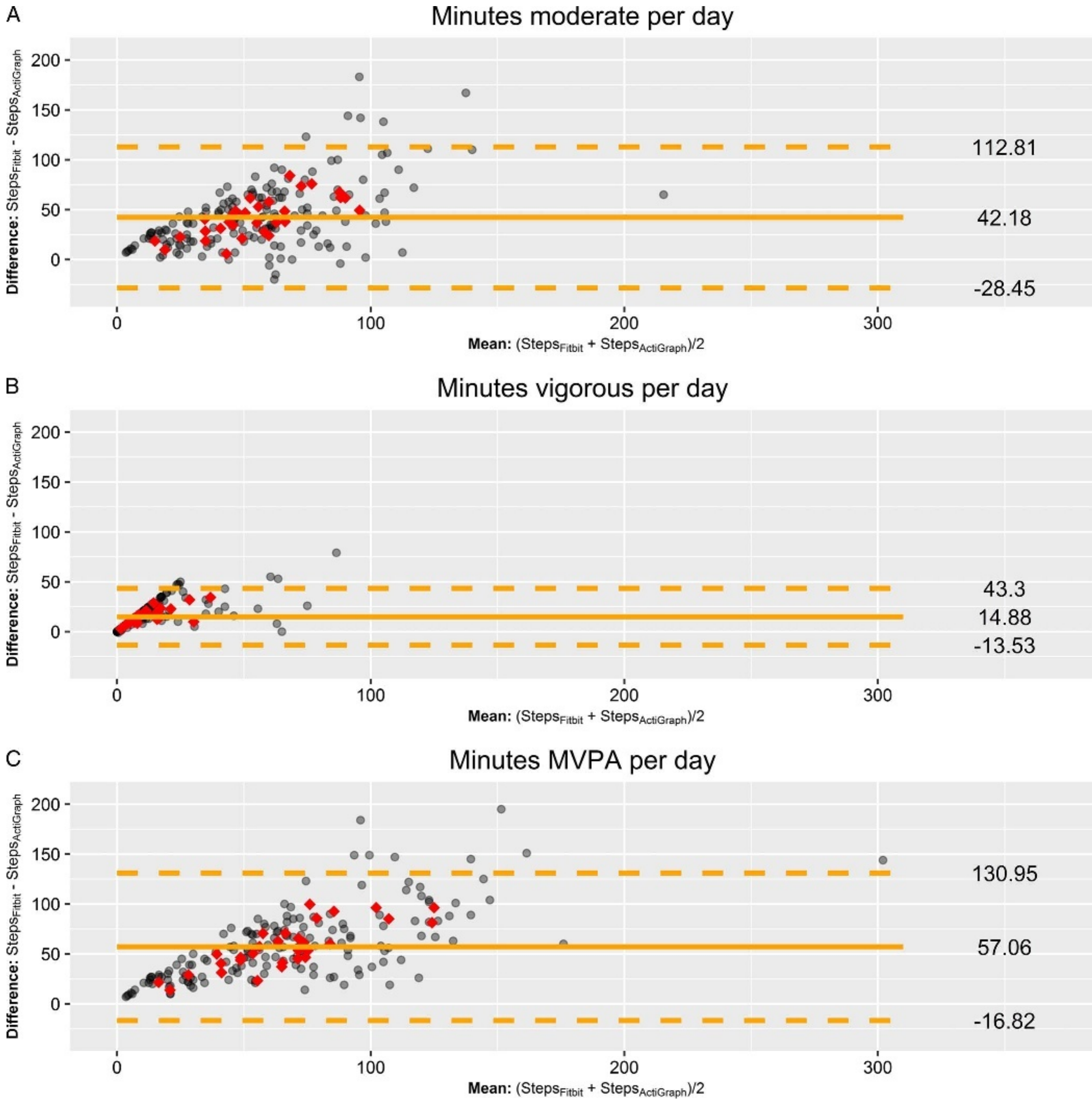

FIGURE 3—Bland–Altman plot for minutes of moderate physical activity, vigorous physical activity, and moderate to vigorous physical activity (MVPA) per day, for measurements with the ActiGraph and Fitbit, with adjusted means and LoA. A, Bland–Altman plot for minutes of moderate physical activity per day. B, Bland–Altman plot for minutes of vigorous physical activity per day. C, Bland–Altman plot for minutes of moderate to vigorous physical activity per day. The *x*-axis shows the average of the two measurements, and the *y*-axis shows the differences between the two methods. The mean values of all steps per day, hour, and minute per individual are plotted in *red*.

accelerometer, which are mentioned previously, should be taken into account. Second, we used Troiano's cut points to categorize time spent in different intensities what may have influenced the results: the cut points are based on regression models and do not adjust for the individual's characteristics and Fitbit's algorithm to categorize time spent at different intensities are unknown and most likely differ from Troiano's cut points. Troiano's cut points (or any other ActiGraph cut points) are certainly not a gold standard for determining physical activity and have their own shortcomings. This is also illustrated by the results of Lee et al. (16) discussed previously that indicated the Fitbit One might be more suitable for the assessment of energy expenditure than the ActiGraph when using the traditional linear regression models. Third, because of dichotomous data for the intensity levels per minute, it was not possible to assess the validity for minutes of moderate, vigorous, and moderate to vigorous physical activity per minute. Lastly, the current study assessed the validity of the Fitbit while worn on the hip. Therefore, the current study cannot generalize the findings to situations in which the Fitbit is worn on other sides (e.g., wrist or bra).

In conclusion, this is the first study examining the validity of the Fitbit One for the assessment of levels of physical activity in real-life using smaller time intervals, which is essential for high quality self-monitoring, which is increasingly used in physical activity intervention studies. Although the Fitbit One overestimates the step activity somewhat, the Fitbit One can be considered a valid device to assess step activity even for real-time minute-by-minute self-monitoring. However, when being compared with the ActiGraph, the Fitbit One appears less suitable to assess time spent in moderate activity, vigorous activity, and moderate–vigorous physical activity. To date, these data are not shown to the user and validity should be improved before this should be done. Further research is needed to examine whether the Fitbit One can be used as a (scientific) measurement tool. Thus, on the basis of our findings, the Fitbit One appears as suitable to deliver real-time feedback based on step data for self-monitoring, even for small time intervals as minute and hour intervals. Furthermore, the Fitbit One also appears as a valid instrument to assess daily step activity.

The authors thank all participants who were willing to participate in our study. They especially want to thank Julienka Mollee for her contribution during the data collection.

This research is supported by Philips and Technology Foundation STW, Nationaal Initiatief Hersenen en Cognitie NIHC, under the Partnership program Healthy Lifestyle Solutions (grant no. 12014).

A. V. H. is a paid member of Philips Research. For the remaining authors, none have a professional relationship with companies or manufacturers who might benefit from the results of the present study.

The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports Medicine.

Author's Contribution: A. M. collected the data, preformed the analyses, drafted the manuscript, and incorporated all feedback. H. P. V. D. P. provided intellectual input and feedback and approved the final version of the manuscript. A. V. H. provided intellectual input and feedback, supported in the analyses, and approved the final version of the manuscript. J. W. R. T. advised on the analyses, provided intellectual input and feedback, and approved the final version of the manuscript. J. B. provided intellectual input and feedback and approved the final version of the manuscript. S. J. T. V. provided intellectual input and feedback and approved the final version of the manuscript.

REFERENCES

1. Bassett DR Jr, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc*. 2000;32(9 Suppl): S471–80.
2. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a fitbit-based physical activity intervention for women. *Am J Prev Med*. 2015;49(3):414–8.
3. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*. 2015;313(6):625–6.
4. Chau JY, van der Ploeg HP, Dunn S, Kurko J, Bauman AE. A tool for measuring workers' sitting time by domain: the Workforce Sitting Questionnaire. *Br J Sports Med*. 2011;45(15):1216–22.
5. Chau JY, Van Der Ploeg HP, Dunn S, Kurko J, Bauman AE. Validity of the occupational sitting and physical activity questionnaire. *Med Sci Sports Exerc*. 2012;44(1):118–25.
6. Cust AE, Smith BJ, Chau J, et al. Validity and repeatability of the EPIC physical activity questionnaire: a validation study using accelerometers as an objective measure. *Int J Behav Nutr Phys Act*. 2008;5:33.
7. Dontje ML, de GM, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the Fitbit activity tracker: an inter-device reliability study. *J Med Eng Technol*. 2015;39(5):286–90.
8. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12(1):159.
9. Evenson KR, Terry JW Jr. Assessment of differing definitions of accelerometer nonwear time. *Res Q Exerc Sport*. 2009;80(2):355–62.
10. 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. *Int J Behav Nutr Phys Act*. 2015;12:42.
11. Fitbit Inc. Fitbit. 2015. 26-3-2015. Available from: http://www.fitbit.com.
12. Fitbit Inc. How does my tracker count steps? 2015. 13-7-2015. Available from: http://help.fitbit.com/articles/en_US/Help_article/How-does-my-tracker-count-steps.
13. Fitbit Inc. What are very active minutes? 2015. 23-12-2015. Available from: https://help.fitbit.com/articles/en_US/Help_article/What-are-very-active-minutes.
14. 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. *J Med Internet Res*. 2016;18(9):e239.
15. Hallal PC, Reichert FF, Clark VL, et al. Energy expenditure compared to physical activity measured by accelerometry and self-report in adolescents: a validation study. *PLoS One*. 2013; 8(11):e77036.
16. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc*. 2014;46(9):1840–8.
17. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *J Med Internet Res*. 2014;16(8):e192.
18. Mackinlay M. Phases of accuracy diagnosis: (in)visibility of system status in the Fitbit. *Intersect*. 2013;6(2):1–9.
19. Mâsse LC, Fuemmeler BF, Anderson CB, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc*. 2005;37(11 Suppl): S544–54.
20. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Med Sci Sports Exerc*. 2016;48(3):457–65.
21. Simpson LA, Eng JJ, Klassen TD, et al. Capturing step counts at slow walking speeds in older adults: comparison of ankle and waist placement of measuring device. *J Rehabil Med*. 2015;47(9):830–5.

22. Storm FA, Heller BW, Mazza C. Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS One*. 2015;10(3):e0118723.
23. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport*. 2014;17(5):496–500.
24. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.
25. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. *Can J Public Health*. 2006;97(1 Suppl):S10-5–10-6.
26. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip physical activity monitor as a measure of free-living physical activity. *BMC Res Notes*. 2014;7:952.
27. Vooijs M, Alpay LL, Snoeck-Stroband JB, et al. Validity and usability of low-cost accelerometers for Internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. *Interact J Med Res*. 2014;3(4):e14.
28. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity—a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17(2):127–39.