


# Fitbit's accuracy to measure short bouts of stepping and sedentary behaviour: validation, sensitivity and specificity study

DIGITAL HEALTH  
Volume 10: 1–16  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241262710  
journals.sagepub.com/home/dhj



Julie Delobelle<sup>1,2</sup> , Elien Lebuf<sup>1,2</sup>, Delfien Van Dyck<sup>1</sup>, Sofie Compennolle<sup>1,2</sup>, Michael Janek<sup>3</sup>, Femke De Backere<sup>4</sup> and Tomas Vetrovsky<sup>3</sup>

## Abstract

**Objective:** This study aims to assess the suitability of Fitbit devices for real-time physical activity (PA) and sedentary behaviour (SB) monitoring in the context of just-in-time adaptive interventions (JITAs) and event-based ecological momentary assessment (EMA) studies.

**Methods:** Thirty-seven adults (18–65 years) and 32 older adults (65+) from Belgium and the Czech Republic wore four devices simultaneously for 3 days: two Fitbit models on the wrist, an ActiGraph GT3X+ at the hip and an ActivPAL at the thigh. Accuracy measures included mean (absolute) error and mean (absolute) percentage error. Concurrent validity was assessed using Lin's concordance correlation coefficient and Bland–Altman analyses. Fitbit's sensitivity and specificity for detecting stepping events across different thresholds and durations were calculated compared to ActiGraph, while ROC curve analyses identified optimal Fitbit thresholds for detecting sedentary events according to ActivPAL.

**Results:** Fitbits demonstrated validity in measuring steps on a short time scale compared to ActiGraph. Except for stepping above 120 steps/min in older adults, both Fitbit models detected stepping bouts in adults and older adults with sensitivities and specificities exceeding 87% and 97%, respectively. Optimal cut-off values for identifying prolonged sitting bouts achieved sensitivities and specificities greater than 93% and 89%, respectively.

**Conclusions:** This study provides practical insights into using Fitbit devices in JITAs and event-based EMA studies among adults and older adults. Fitbits' reasonable accuracy in detecting short bouts of stepping and SB makes them suitable for triggering JITA prompts or EMA questionnaires following a PA or SB event of interest.

## Keywords

physical activity, sedentary behaviour, event-based ecological momentary assessment (EMA), ecological momentary assessment, JITA, just-in-time adaptive intervention, accelerometry, Fitbit, wearable, step

Submission date: 5 February 2024; Acceptance date: 1 June 2024

## Introduction

Physical inactivity and sedentary behaviour have unfavourable effects on mental and physical well-being.<sup>1–3</sup> However, many people, especially older adults, are insufficiently active, making physical inactivity and sedentary time the leading risk factors for noncommunicable diseases and death worldwide.<sup>1,4</sup> Health behaviour interventions have been developed to promote physical activity (PA) and reduce sedentary behaviour (SB), but they mostly

<sup>1</sup>Physical Activity & Health, Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

<sup>2</sup>Research Foundation Flanders (FWO), Brussels, Belgium

<sup>3</sup>Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic

<sup>4</sup>Faculty of Engineering and Architecture, Department of Information Technology, Ghent University, Ghent, Belgium

### Corresponding author:

Julie Delobelle, Physical Activity & Health, Department of Movement and Sports Sciences, Ghent University, Watersportlaan 2, 9000 Ghent, Belgium.  
Email: Julie.Delobelle@UGent.be



achieve limited and short-term behaviour change.<sup>5–7</sup> The lack of long-term effectiveness might (partly) be explained by the fact that current interventions assume determinants of movement behaviours to be static (i.e. stable over time) and neglect their fluctuations over time.<sup>8,9</sup> However, previous research indicated that PA and SB are dynamic processes influenced by momentary and daily fluctuations in determinants.<sup>10</sup> For example, a recent study in older adults showed that multiple individual-level determinants (emotions, physical complaints, intention and self-efficacy) of older adults could vary over time.<sup>11</sup> Besides, individuals are constantly exposed to social and physical environmental factors, as well as events throughout the day, which might influence a person's behaviour. Hence, determinants of movement behaviours can vary within individuals over time (within and between days), which suggests the need for a dynamic approach in interventions targeting PA and SB.<sup>12</sup>

Just-in-time adaptive interventions (JITAI) are intervention designs that address the dynamic nature of the determinants of movement behaviours (e.g. engagement in PA and reduction of SB) by providing the right type and amount of support at the right time.<sup>13,14</sup> Specifically, these interventions adapt in real time to an individual's changing internal (e.g. when an individual is most likely to be receptive and needs it most) and external (e.g. social and environmental context) states.<sup>15</sup> Specifically, JITAI can respond effectively when individual-level determinants (e.g. self-efficacy) are at optimal levels by identifying fluctuations in these determinants before prompting the intervention. In addition, prompting may also be contingent on behaviour. For example, an encouraging prompt to prolong a walk can be triggered on a participant's smartphone when the participant is walking from work in the afternoon, or a prompt suggesting a short bout of exercise can be triggered after a prolonged episode of sitting and watching TV. Despite JITAI's potential to promote a physically active lifestyle, their development is still in its early stages.<sup>14</sup>

Developing effective JITAI requires better insight into the dynamic determinants of PA and SB. To capture the time- and context-dependent variation of behavioural determinants, researchers commonly use ecological momentary assessment (EMA). EMA involves repeated sampling of behaviours and experiences during people's everyday lives, thus maximising ecological validity and minimising recall bias.<sup>16–18</sup> Within an EMA study, individuals are usually prompted at fixed or randomly allocated times during the day (i.e. time-based EMA).<sup>19</sup> By doing so, many events of interest might be missed. For example, the probability of prompting a questionnaire during a bout of PA will be small, resulting in a lack of important information about why someone is engaging in PA at a specific time and context. Since increasing prompt frequency would result in a higher user burden, event-based EMA is an

innovative data collection method to assess experiences, feelings and contexts during or following a specific event (e.g. a short bout of PA or prolonged episode of sitting). Although event-based EMA is promising for identifying the individual and environmental dynamic determinants of PA and SB, its use in this field of research is still in its infancy.

In both JITAI and event-based EMA research, the continuous monitoring of participants' behaviour is needed to enable the detection of the events of interest (e.g. 5 min of sustained walking or 30 min of prolonged sitting). Although consumer-based activity trackers like Fitbit are user-friendly devices with the potential to be used as a monitoring tool to trigger both JITAI prompts and EMA questionnaires,<sup>20</sup> it is still not known whether they can accurately measure short bouts of PA and SB. The Fitbit already appeared to be nonvalid for giving real-time moderate to vigorous physical activity (MVPA) feedback on a 15-min level.<sup>21</sup> However, for steps, Fitbit was identified as a valid measurement tool on a 15-min level compared to the ActiGraph GT3X+.<sup>21</sup> Nevertheless, to provide real-time feedback on minutes of (in)activity or to serve as a wearable sensor for event-based EMA and JITAI, it is important to assess its validity in measuring stepping and SB on even a smaller time scale (e.g. 1, 5 and 10 min). This would enable researchers to assess individual and environmental determinants immediately after a short bout of walking and SB or to provide tailored support at the right time and context using JITAI.<sup>22</sup>

To provide a rigorous and transparent validation of Fitbits for use in JITAI and event-based EMA and to enable researchers to make informed decisions on what device to use in a particular study, it is recommended to adopt a standardised validation framework.<sup>23</sup> Keadle et al. introduced a four-stage process framework designed to facilitate the development and validation of measures of physical behaviour.<sup>24</sup> The framework progresses from mechanical (phase 0) and calibration testing (phase 1), through validation in a controlled semi-structured laboratory (phase 2) and naturalistic setting (phase 3), to implementing the new device or method (phase 4).<sup>24</sup> Fitbit devices routinely undergo mechanical and calibration testing to ensure that they meet the manufacturer's quality standards and that they are accurate and reliable for consumers.<sup>25</sup> Furthermore, a body of literature has indicated that Fitbits' accuracy in measuring steps in a controlled semi-structured laboratory setting is acceptable.<sup>26</sup> However, studies in a naturalistic setting (phase 3) validating Fitbits for detecting short episodes of stepping and sitting are lacking.

## Objectives

The objective of this study was to conduct a phase-3 validation of Fitbits<sup>24</sup> and provide practical considerations on using Fitbits for JITAI and event-based EMA studies in

adults (18–65 years) and older adults (65+). Specifically, we aimed (1) to explore the concurrent validity of two models of Fitbits with ActiGraph to ensure they can accurately capture short bouts of stepping and (2) to examine the sensitivity and specificity of Fitbits to detect short bouts of stepping (1, 5, 10 and 20 min) and SB (1, 20, 30 and 60 min) using various steps-per-minute thresholds. For stepping, this allows researchers to get an idea about the accuracy of Fitbits to detect the episodes of walking of various durations (1 to 20 min) and walking cadence (60 to 120 steps per minute) for the age group they are interested in. For SB, we calculated the optimal threshold (in steps per minute) of Fitbits to capture an event of prolonged sitting (1 to 60 min) measured with ActivPAL in adults and older adults. In addition, we investigated the day-level concurrent validity and inter-device agreement between two different Fitbit models to find out if these results can be generalised to other Fitbit models.

## Methods

This observational study took place in Belgium and the Czech Republic from May to December 2021. Over a 3-day period, participants simultaneously wore four different devices: two models of Fitbit on the wrist, an ActiGraph GT3X+ on the hip and an ActivPAL on the thigh.

### Participant recruitment

A convenience sample of 37 adults (aged between 18 and 65 years) and 32 older adults (aged over 65 years) was recruited from the researchers' social network, including friends, family and colleagues. Only participants who were able to walk at least 100 m independently were included in the study. There were no other eligibility criteria for participation in the study. The participants were recruited, and data were collected in two different countries (Belgium (Flanders,  $n=40$ ) and the Czech Republic (Prague,  $n=29$ )). Participants received an informative letter outlining the study's objectives, design, purpose, data confidentiality and their right to withdraw from the study at any time without the need to provide a reason. All participants read and signed the informed consent form before inclusion in the study. The study protocol was approved by the Ethics Committee of Ghent University (BC-09448) and Charles University (No. 299/2021).

### Measures

We investigated the concurrent validity of two wearables (Fitbit Ionic and Fitbit Inspire 2) using two triaxial accelerometers (ActiGraph GT3X+ and ActivPAL4) as a reference. The Fitbit Ionic and Inspire 2 (Fitbit Inc., San Francisco, CA) convert motion patterns to step counts in 1-min epochs. The Inspire 2 is an activity tracker, which is specifically designed to monitor health

and activity. The Ionic is a smartwatch, which, besides tracking activity levels, also features other functions (e.g. receiving calls or delivering notifications). To assess the accuracy of the Fitbits in measuring PA, the hip-worn ActiGraph GT3X+ (ActiGraph, Pensicola, FL, USA) was used as a reference device. This accelerometer was found reliable and valid for measuring steps in different settings.<sup>27,28</sup> In addition, SB was assessed using the thigh-worn ActivPAL4 (Pal Technologies Ltd., Glasgow, UK) chosen for its superior accuracy in detecting SB compared to the ActiGraph GT3X+. Positioning the device on the thigh allows for a more accurate detection of posture changes, including sitting, standing and lying down. ActivPAL4 has proven to be a reliable and valid accelerometer for assessing SB in free-living conditions<sup>29–32</sup> and in laboratory settings.<sup>33</sup>

### Procedures

Prior to data collection, all participants were either invited to the lab or visited at home to sign the informed consent, hand over the four devices and administer a short questionnaire on socio-demographic variables, such as age, gender, height, weight and nationality (see Supplementary Materials 1). Participants were asked to wear the ActiGraph and ActivPAL accelerometers and the two different Fitbit models (Ionic and Inspire 2) concurrently for three consecutive days while maintaining their normal behaviour. The Fitbit wearables were worn simultaneously next to each other at the non-dominant wrist in a randomly assigned order to eliminate the potential effect of where the Fitbits were worn (e.g. Ionic proximal and Inspire 2 distal or vice versa). The two Fitbit models logged the users' minute-by-minute step count and heart rate. Since the data collection period only lasted 3 days, no smartphone was required to synchronise them with the Fitbit app as their internal memory can store up to 7 days of data. ActivPAL was made waterproof and attached by a hypoallergenic adhesive tape to the midpoint of the right upper thigh. ActiGraph was fitted to the participants' right hip and was the only device that was removed during sleep time (to increase participants' comfort), showering or water activities (because it is not waterproof). Both devices were placed at their respective positions in accordance with the manufacturer's instructions and previous research. To ensure that the accelerometers' and the Fitbits' internal clocks were synchronised at the exact same time, all the measurement instruments were initialised using the same laptop.<sup>34</sup> Immediately after the testing period, the devices were collected, and the Fitbits were synchronised with the Fitbit application to prevent data loss.

### Data processing

Step count (minute-by-minute) and heart rate data (second-by-second) were extracted from the Fitbit accounts with an application programming interface (API) using the OAuth

2.0 Client Library and the Fitbitr package in R.<sup>35</sup> Rolling (moving) averages of steps were calculated for bouts of various durations (5, 10 and 20 min).

ActiGraph data were initialised, downloaded and processed using ActiLife version 6.13.4 software (ActiGraph, Fort Walton Beach, Florida, USA). ActiGraph data were recorded at a sampling frequency of 30 Hz and converted into 60-s epochs without a low-frequency extension (LFE) filter. Troiano's (2007) algorithm was used to define non-wear time, which was defined as a minimum of 60 min of 0 counts per minute (cpm) with an allowance of 2 min of interruptions.<sup>36</sup> Minute-by-minute step data were extracted from ActiLife and rolling averages were calculated for 5, 10 and 20 min.

Events of sitting, standing, stepping, cycling, sleeping, lying and seated traveling were classified using ActivPALs proprietary software. Data were exported from ActivPAL into the event format (e.g. format describing the sequence of events with the corresponding time stamps of their start and end) and converted into second-by-second data. Then, second-by-second ActivPAL data were converted to 60-s (minute) epoch data. Minutes consisting exclusively of 60 consecutive seconds of SB, such as sitting, travelling in a seated position and lying (classified using ActivPAL proprietary software) were considered sedentary minutes. Minutes in which SB was interrupted (if only for 1 s) were considered non-sedentary minutes. Hence, zero tolerance for other behaviours (e.g. cycling and stepping) was allowed for sedentary minutes. Moreover, all sedentary bouts (20, 30 and 60 min) consisted only of these sedentary minutes. Sleeping minutes were excluded from the analysis.

Only minutes that fulfilled all these criteria were included in the analyses: (a) time of the day between 9:00 a.m. and 8:59 p.m.; (b) ActiGraph data flagged as "wear" when applying the Troiano algorithm<sup>36</sup>; and (c) both Fitbits recorded at least one heart rate measurement as a marker that the devices were actually worn. All other minutes were excluded from the analysis to ensure that we only analysed minutes when all four devices were worn concurrently.

### Statistical analysis

Analyses were performed using R (version 4.2.0). First, we calculated several commonly used measures of accuracy,<sup>26,37,38</sup> including mean error (ME), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE), for the recorded step count between ActiGraph and each of the Fitbit devices on a daily level. These measures provide quantitative insights into the degree of agreement or discrepancy between the step counts recorded by ActiGraph and those recorded by the Fitbit devices. The percentage error was calculated as a difference between the Fitbit data and the ActiGraph data divided by the ActiGraph data. The percentage errors and their absolute values were averaged to compute the mean percentage error (MPE) and the mean absolute

percentage error (MAPE), respectively. The MPE value assessed the degree of the overall overestimation or underestimation of the Fitbit against the ActiGraph, whereas the MAPE value provided the most relevant and comparable indicator of individual error because it accounted for both overestimation and underestimation. Second, to assess the concurrent validity of Fitbits to measure steps in short bouts, we calculated the ME, MAE and Lin's concordance correlation coefficients (CCC) and constructed Bland-Altman plots<sup>34,38,39</sup> for 1, 5, 10 and 20 consecutive minutes. In these analyses, we only included minutes where the ActiGraph recorded  $\geq 60$  steps (considered as sustained stepping).<sup>40</sup> For example, for 10 consecutive minutes in which ActiGraph detected more than 60 steps, we obtained five valid 5-min bouts (i.e. 0-5, 1-6, 2-7, 3-8, 4-9 and 5-10). To assess the degree of agreement, the following strength-of-agreement criteria were applied:  $<0.90$  poor,  $0.90$  to  $0.95$  moderate,  $0.95$  to  $0.99$  substantial, and  $>0.99$  almost perfect.<sup>39</sup> Third, we examined the potential effect of the measurement day and site (i.e. Belgium and the Czech Republic) on the minute-level difference using a linear regression model. Fourth, to explore the accuracy of Fitbits in detecting stepping events, we calculated sensitivity and specificity for different thresholds (60, 80, 100 and 120 steps/min) and various lengths (1, 5, 10 and 20 min) of stepping events as detected by ActiGraph. In addition, Cohen's Kappa was calculated to evaluate the agreement between Fitbits and ActiGraph in classifying individual bouts. The following degree-of-agreement criteria were applied:  $0-0.20$ : slight agreement;  $0.21-0.40$ : fair agreement;  $0.41-0.60$ : moderate agreement;  $0.61-0.80$ : substantial agreement;  $0.81-0.99$ : near perfect agreement; and  $1$ : perfect agreement.<sup>41</sup> Fifth, to define the optimal Fitbit threshold in steps/min for detecting sedentary events as identified by the ActivPAL, we performed receiver operating characteristic (ROC) curve analyses and calculated the sensitivity and specificity of this threshold to detect sedentary events of various durations (1, 20, 30 and 60 min). The optimal cut-off was identified using Youden's J statistic<sup>42</sup> as a point on the ROC curve with maximum distance to the identity (diagonal) line. We calculated Cohen's Kappa to evaluate the agreement between Fitbits and ActivPAL in detecting individual bouts. Finally, to evaluate the inter-device agreement between the two Fitbit models (Ionic and Inspire 2) on a daily level, we calculated the ME and MAPE. Moreover, to assess the inter-device agreement for short bouts, we calculated the ME, MAE and CCC and constructed Bland-Altman plots for 1, 5, 10 and 20 consecutive minutes.

## Results

### Participant characteristics

In total, 29 participants were recruited in the Czech Republic and 40 in Belgium. All 69 participants wore the

four devices for three consecutive days. Six participants were excluded from the analysis due to technical issues resulting in a complete lack of Fitbit-recorded heart rate or steps or ActivPAL data. Thus, all analyses were performed on the remaining 63 participants (33 adults and 30 older adults). Adult participants had a mean age of  $31.76 \pm 17.77$  years and a mean body mass index (BMI) of  $22.81 \pm 3.19$  kg/m<sup>2</sup>. Almost half of them were male (41%). Older adult participants had a mean age of  $76.93 \pm 8.40$  years and a BMI of  $24.85 \pm 4.25$  kg/m<sup>2</sup>. Only 25% of them were male.

### Concurrent validity

**Day level.** Table 1 presents the mean daily steps recorded by the Fitbit Inspire 2, Ionic and ActiGraph accelerometer. The ActiGraph-recorded step count per day ranged between 992 and 25,060 in adults and between 295 and 26,400 in older adults. Both Fitbit devices overestimated the number of steps per day in both age groups compared to the ActiGraph.

**1-, 5, 10- and 20-min levels.** In Table 2, the ME and the MAE in step counts are shown for minutes with ActiGraph-recorded steps  $\geq 60$ . The results are provided separately for both Fitbit models, each age group and four bout durations (1, 5, 10 and 20 min). The results show that in adults, both Fitbit models overestimated the step counts for the 1-, 5-, 10- and 20-min bouts. However, this overestimation decreased with longer bout duration. The opposite effect was found in older adults since an underestimation of the Fitbits was found on the 5-, 10- and 20-min levels. This effect increased with longer bout duration.

Bland–Altman plots were used to plot the differences in step count between the ActiGraph and the Fitbit devices (y-axis) against the mean step count of the two measuring instruments (x-axis). Since the difference was calculated as Fitbit minus ActiGraph, a positive value of the difference indicates an overestimation by Fitbit, whereas a negative value indicates an underestimation. Perfect agreement is indicated by a mean difference of zero, suggesting that there is no systematic bias between the two methods. The range between the upper and lower limits of agreement includes 95% of differences between the two devices and reflects the accuracy of the Fitbits to measure steps. Figures 1 (adults) and 2 (older adults) present the Bland–Altman plots and the corresponding CCCs on the 1-, 5-, 10- and 20-min levels for each Fitbit device against the ActiGraph. The sharp cut-off on the left side of 1-min level plots is caused by the exclusion of steps below 60 per minute. In adults, these plots show an improvement in accuracy with increasing cadence. In older adults, the Bland–Altman plots show an underestimation of steps with increasing walking pace. The correlations between the ActiGraph and Fitbits are poor to almost perfect,

ranging from 0.80 to 0.99 in adults and 0.85 to 0.90 in older adults.

**Influence of time and site.** Multiple linear regression analyses were used to test the potential effect of the day of measurement and site of participant recruitment (Belgium vs. the Czech Republic) on the minute-level difference in step count between the Fitbit devices and ActiGraph. The day of measurement significantly affected the minute-level difference for Inspire 2 ( $\beta = 0.12$ ,  $SE = 0.04$ ,  $P < .004$ ) and Ionic ( $\beta = 0.17$ ,  $SE = 0.04$ ,  $P < .001$ ). Specifically, for each consecutive day of measurement, the minute-level difference between ActiGraph and Fitbit got greater by 0.12 steps for the Inspire 2 and 0.17 steps for the Ionic. No significant effect of the site (i.e. Belgium or the Czech Republic) was found on the minute-level difference in steps for any of the Fitbit devices ( $\beta = -0.05$ ,  $SE = 0.07$ ,  $P = .45$  for the Inspire 2 and  $\beta = -0.002$ ,  $SE = 0.07$ ,  $P = .98$  for the Ionic).

### Accuracy in detecting stepping and sedentary events

**Stepping events.** Tables 3 and 4 present an overview of Fitbit's accuracy in capturing short bouts of stepping for several step-rate thresholds for both adults and older adults. Specifically, we report sensitivities and specificities of the Fitbits in detecting stepping events above different cut-off values (60, 80, 100 and 120 steps/min) on a short time scale (1, 5, 10 and 20 min). Sensitivities ranged between 87% and 100% in adults and between 34% and 100% in older adults for several thresholds, indicating a good sensitivity of Fitbits in detecting short bouts of stepping in all cases, except for stepping above 120 steps/min in older adults. Furthermore, specificities for each threshold and duration were greater than 97% in both adults and older adults. For example, if researchers use the Fitbit Ionic to elicit an EMA questionnaire for adults after a 5-min stepping event, in which the participant takes at least on average  $\geq 60$  steps/min, they can expect a sensitivity of 95.72% and a specificity of 98.26%. Similarly, interventionists who wish to use Fitbit Inspire 2 to stimulate older adults to remain physically active after a 10-min bout of  $\geq 80$  steps/min on average can expect a sensitivity and a specificity of 98.01% and 99.09%, respectively. In other words, in approximately 98 out of 100 stepping bouts, the bout will be captured correctly, and two will be missed. Moreover, in approximately one out of 100 non-stepping bouts, it will be falsely detected as a stepping bout. In addition, except for stepping above 120 steps/min in older adults, Cohen's Kappa's ranged between 0.79 and 0.91, indicating substantial to nearly perfect agreement between Fitbits and ActiGraph.

**Sedentary events.** Additionally, we provide specific cut-off values (in steps/min) that result in optimal sensitivities and specificities of Fitbits to capture bouts of SB of different lengths (20, 30 and 60 min) as identified by ActivPAL.

**Table 1.** Mean steps per day measured concurrently by Fitbit Inspire 2, Ionic and ActiGraph for adults and older adults. The mean error (ME), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE) of Fitbits (with ActiGraph as a reference device) in the mean steps are presented.

Accelerometer	Mean (SD)	ME <sup>a</sup> (SD)	MAE <sup>b</sup> (SD)	MPE <sup>c</sup>	MAPE <sup>d</sup>
<b>Adults</b>					
ActiGraph	7161 (4587)				
Ionic	8553 (5413)	1392 (1688)	1471 (1619)	19.44	20.54
Inspire 2	8970 (5590)	1810 (1870)	1856 (1823)	25.27	25.92
<b>Older adults</b>					
ActiGraph	6658 (5396)				
Ionic	8116 (5484)	1458 (1926)	1707 (1704)	21.90	25.65
Inspire 2	8545 (5641)	1887 (1877)	1954(1806)	28.35	29.35

<sup>a</sup>ME: mean error, <sup>b</sup>MAE: mean absolute error, <sup>c</sup>MPE: mean percentage error, <sup>d</sup>MAPE: mean absolute percentage error.

**Table 2.** Mean difference in steps per minute between Fitbit Inspire 2, Ionic and ActiGraph.

Duration (min)	Ionic vs ActiGraph	Inspire 2 vs ActiGraph
	ME <sup>a</sup> (MAE <sup>b</sup> ) in steps	ME <sup>a</sup> (MAE <sup>b</sup> ) in steps
<b>Adults</b>		
1	1.09 (8.90)	2.05 (9.06)
5	1.60 (3.76)	1.92 (3.80)
10	1.07 (2.20)	1.34 (2.32)
20	0.36 (1.28)	0.51 (1.32)
<b>Older adults</b>		
1	0.19 (7.41)	0.69 (7.25)
5	-1.08 (5.52)	-0.56 (5.17)
10	-1.88 (5.73)	-1.33 (5.35)
20	-2.72 (6.51)	-1.90 (5.80)

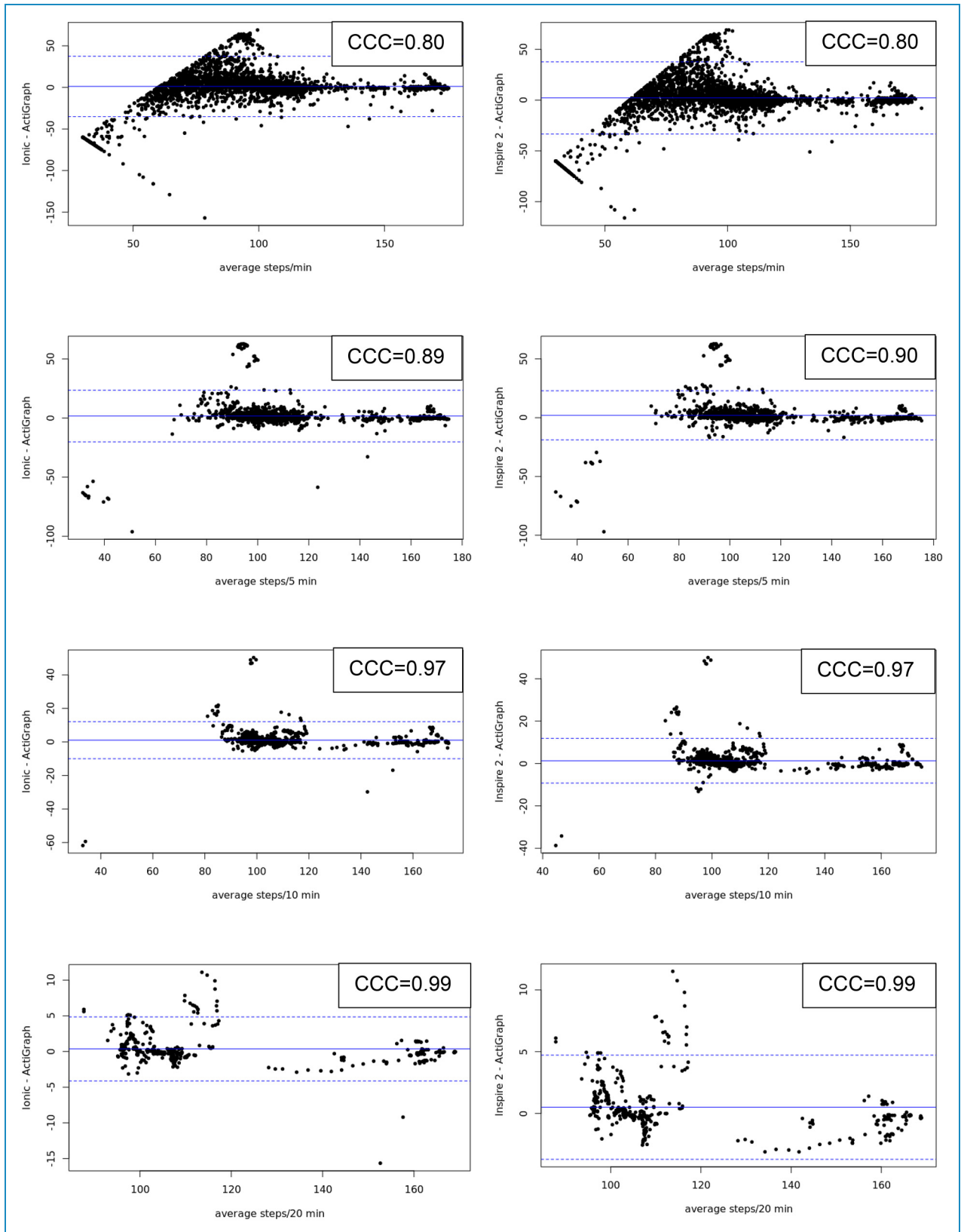
<sup>a</sup>ME: mean error, <sup>b</sup>MAE: mean absolute error.

The optimal cut-off values and their respective sensitivities and specificities were obtained from ROC curves and are presented in Table 5. Sensitivities ranged between 93% and 100% and specificities between 89% and 100%. Only during 1-min epochs, Fitbits obtained low specificities

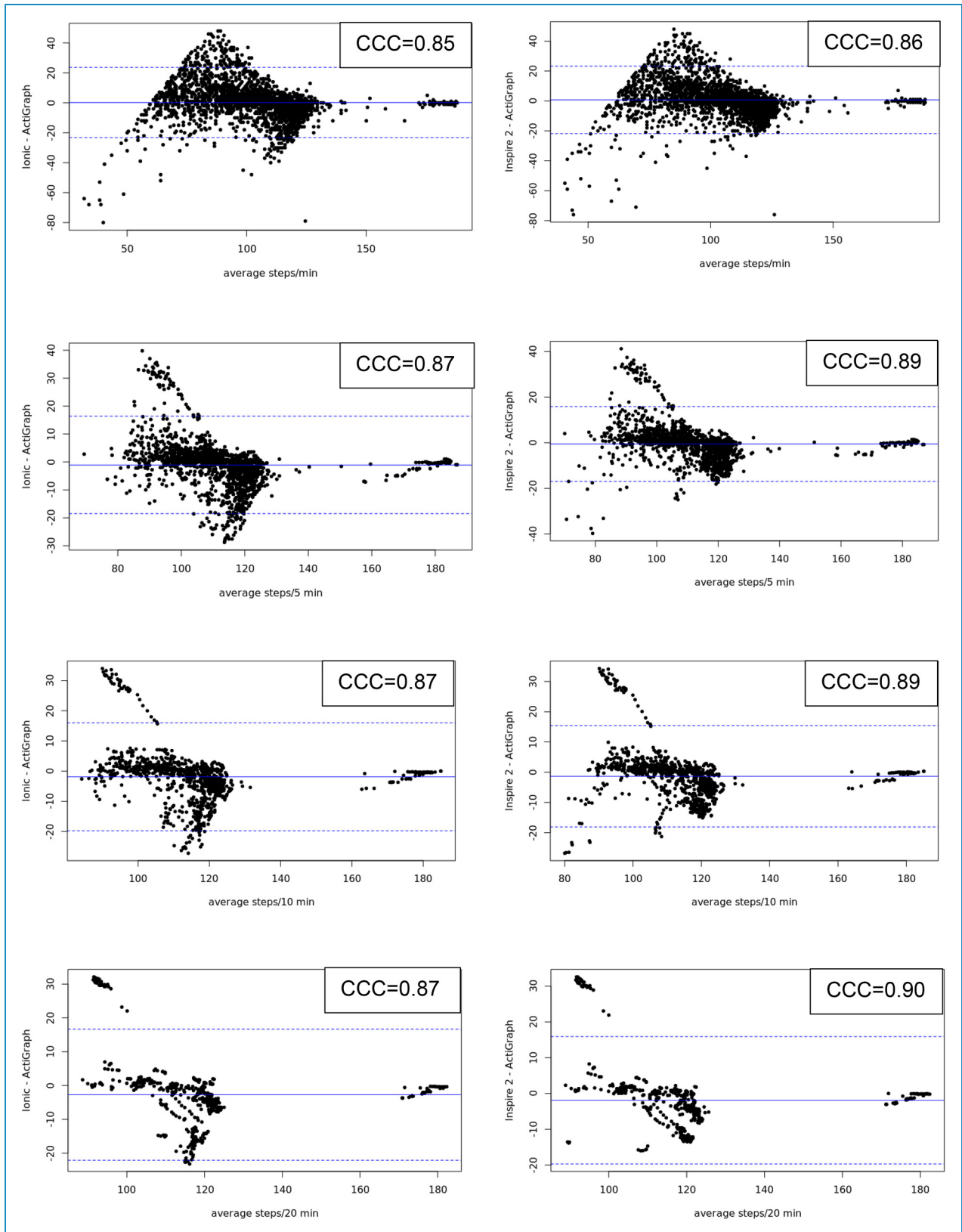
(61%–71%). In addition, the Cohen's Kappa's indicated substantial to nearly perfect agreement between Fitbits and ActivPAL in detecting sedentary events of interest (20, 30 and 60 min) as they ranged between 0.77 and 0.94. Table 5 can be interpreted as follows: a sedentary event of 30 min in older adults was best captured by a cut-off value of smaller or equal to 2.08 average steps/min, resulting in a sensitivity of 93.27% and a specificity of 91.99% with Fitbit Inspire 2. This means that in approximately 93 out of 100 true sedentary bouts, the bout will be captured correctly, and seven bouts will be missed. Moreover, in eight out of 100 true non-sedentary bouts, a bout will be falsely captured as a sedentary bout. The corresponding ROC curve can be found in Figure 3.

### Inter-device agreement between the two Fitbits

Overall, Inspire 2 recorded more steps per day compared to Ionic in adults (mean difference 415, SD 489) and in older adults (mean difference 429, SD 540). Moreover, we observed minimal differences in step counts between the two Fitbit devices on a 1-, 5-, 10- and 20-min time scale, ranging from 0.15 to 1 steps/min. In addition, the CCC ranged from 0.96 to 0.99 in adults and from 0.95 to 0.99 in older adults, indicating a substantial agreement that improved with longer epochs. The Bland-Altman plots show the relationship between the Fitbit Inspire 2 and the Fitbit Ionic for short bout durations (1, 5, 10 and 20 min) in adults and older adults (Figures S1 and S2 in Supplementary Materials 2). These plots show narrow limits of agreement, indicating a good agreement between Ionic and Inspire 2 for measuring steps.



**Figure 1.** Bland-Altman plots of Fitbit Ionic (on the left) and Inspire 2 (on the right) compared to ActiGraph for adults. The mean difference is shown by the middle line. Positive values indicate an overestimation by Fitbit. The dotted lines represent the limits of agreement.



**Figure 2.** Bland-Altman plots of Fitbit Ionic (on the left) and Inspire 2 (on the right) compared to ActiGraph for older adults. The mean difference is shown by the middle line. Positive values indicate an overestimation by Fitbit. The dotted lines represent the limits of agreement.



**Table 3.** Sensitivity and specificity of Fitbits for stepping events for adults (18–65) as identified by the ActiGraph.

		60 steps/min		80 steps/min		100 steps/min		120 steps/min	
		Ionic	Inspire	Ionic	Inspire	Ionic	Inspire	Ionic	Inspire
1	Sensitivity (%)	93.40	94.60	96.40	97.03	94.56	94.75	87.81	86.85
	Specificity (%)	97.43	97.11	98.50	98.30	99.19	99.04	99.77	99.74
	Cohen's Kappa 95% CI <sup>a</sup>	0.79 [0.79– 0.80]	0.78 [0.77– 0.79]	0.84 [0.83– 0.85]	0.83 [0.82– 0.84]	0.86 [0.85– 0.87]	0.85 [0.84– 0.86]	0.82 [0.79– 0.84]	0.80 [0.77– 0.83]
	Occurrence (n)	3908		2892		2059		517	
	Participants (n)	33		33		33		21	
5	Sensitivity (%)	95.72	96.52	98.79	99.19	97.61	98.15	96.30	96.97
	Specificity (%)	98.26	97.95	99.02	98.82	99.47	99.39	99.84	99.83
	Cohen's Kappa 95% CI	0.82 [0.79– 0.84]	0.80 [0.77– 0.83]	0.82 [0.79– 0.84]	0.80 [0.77– 0.83]	0.82 [0.79– 0.84]	0.80 [0.77– 0.83]	0.82 [0.79– 0.84]	0.80 [0.77– 0.83]
	Occurrence (n)	3248		2227		1296		297	
	Participants (n)	33		32		30		13	
10	Sensitivity (%)	96.32	96.49	98.83	99.31	97.60	98.40	99.18	100
	Specificity (%)	98.55	98.30	99.18	99.03	99.62	99.55	99.87	99.87
	Cohen's Kappa 95% CI	0.85 [0.84– 0.86]	0.84 [0.83– 0.84]	0.88 [0.87– 0.89]	0.86 [0.85– 0.88]	0.88 [0.86– 0.89]	0.86 [0.84– 0.88]	0.86 [0.83– 0.89]	0.87 [0.84– 0.90]
	Occurrence (n)	2938		1885		875		245	
	Participants (n)	32		29		26		7	
20	Sensitivity (%)	96.85	97.21	99.01	99.21	97.85	99.41	97.97	100
	Specificity (%)	98.64	98.46	99.32	99.20	99.78	99.75	99.88	99.88
	Cohen's Kappa 95% CI	0.85 [0.84– 0.86]	0.84 [0.82– 0.85]	0.88 [0.86– 0.89]	0.86 [0.85– 0.87]	0.88 [0.86– 0.90]	0.88 [0.86– 0.90]	0.85 [0.81– 0.88]	0.86 [0.82– 0.89]
	Occurrence (n)	2477		1421		512		197	
	Participants (n)	31		23		14		6	

<sup>a</sup>CI = 95% confidence interval.

## Discussion

This study evaluated the concurrent validity, accuracy and the inter-device agreement of two Fitbit models (i.e. Inspire 2 and Ionic) for measuring short bouts of stepping and SB among free-living adults and older adults. The hip-worn ActiGraph GT3x+ was used as a reference measure to validate the

Fitbit devices for stepping events, and the ActivPAL4 was used to evaluate Fitbit's accuracy in detecting sedentary events based on Fitbit-recorded step count. Moreover, the study suggests employing steps per minute rather than specific intensity levels (e.g. MVPA) considering Fitbit's observed limitations in measuring PA intensity.<sup>21</sup>

## Main findings

In general, our results showed that Fitbits can accurately measure short bouts of stepping and SB in adults and older adults (except for stepping above 120 steps/min). Thus, Fitbits can be used as a wearable sensor to collect real-time information about physical behaviour and identify episodes of short walking or prolonged bouts of sitting. This information can be used to trigger event-based EMA questionnaires designed to gain more insight into the context in which (older) adults find themselves during PA or SB. Moreover, Fitbits can be used in JITAI to trigger provision of support at the “right time”, such as after a prolonged bout of SB with the aim of increasing PA or decreasing SB.

**Stepping.** With the exception of stepping above 120 steps/min in older adults, both Fitbit models detected stepping bouts of various durations in both adults and older adults with sensitivity greater than 87%, indicating that a high number of bouts will correctly be detected. For event-based EMA studies, it may be important to capture as many stepping bouts as possible since stepping bouts may be scarce (especially in older adults). Thus, since we obtained sensitivities greater than 87%, we can conclude that Fitbit is a suitable device for detecting stepping for event-based EMA or JITAI purposes. Obtaining even higher sensitivities for a physically inactive population might be beneficial, but not essential, since the main goal remains to promote PA during everyday life.

However, it should be noted that the sensitivities for the highest threshold (i.e. 120 steps/min) in older adults (but not adults) are poor, ranging from 34% to 52%. It can be concluded that there is an underestimation of the Fitbit-recorded step count in older adults with increasing walking pace, which can also be seen in the Bland–Altman plots. A possible explanation for this observation is that walking bouts at lower stepping rates often represent interrupted walking or walking while doing something else, which is associated with a greater proportion of hand movements to steps. As a result, Fitbits overestimate the step count since these hand movements are incorrectly recognised as steps by Fitbits, but not ActiGraph.<sup>43</sup> In contrast, during walking bouts at higher stepping rates (typically representing continuous harmonious walking without erroneous hand movements<sup>44</sup>), Fitbits may miss some steps (which ActiGraph correctly detects), hence their tendency to underestimate. This effect is especially seen in older adults, who walk with less force. Furthermore, the results indicate that underestimation in older adults is more pronounced for longer durations (10 and 20 min). We can speculate that when people walk for a longer duration, their walking is usually more harmonious, with fewer erroneous hand movements than when they walk for just 1 min. Consequently, during longer stepping bouts, Fitbits tend to underestimate the step count in older adults compared to ActiGraph. Moreover, the observed

discrepancies in the step count accuracy between adults and older adults may be attributed to age-related differences in gait mechanics and movement patterns. These age-related differences in gait mechanics may impact the way steps are detected by wearable devices like Fitbit. The device may struggle to accurately capture each step taken by the older adult due to variations in step length, swing speed and rhythm, leading to potential inaccuracies in the step count estimation.

Furthermore, specificities were all greater than 97% in both adults and older adults. Obtaining high specificities may be even more important than high sensitivities because questionnaires/prompts triggered by falsely detected episodes can be confusing for participants and might induce disengagement. Considering the high specificity rates we observed in this study, Fitbit can be used to detect short stepping bouts for event-based EMA or JITAI purposes. In addition, to account for misspecification and prevent starting a questionnaire if the behaviour was wrongly detected, one could first ask whether a short stepping bout was actually performed. To conclude, with the one exception of the poor sensitivity to detect bouts greater than 120 steps/min in older adults, Fitbits can be used to trigger EMA questionnaires and JITAI prompts following a stepping bout (e.g. 5 min of sustained walking).

**Sedentary behaviour.** Since sedentary bouts are more common than PA bouts, achieving high sensitivity may not be so important. However, limiting wrongly detected sedentary bouts (i.e. achieving high specificity) is important since prolonged bouts of SB are very common, and prompting the participant with EMA questionnaires or JITAI prompts unnecessarily may result in a higher burden, more frustration and more drop-out.<sup>45,46</sup> Considering that all sensitivities and specificities were greater than 93% and 89%, respectively, for events of interest (20 to 60 min), Fitbits may serve as a wearable sensor for event-based EMA and JITAI to detect bouts of SB. Since many guidelines advise to interrupt being sedentary after 30 min,<sup>47,48</sup> interventions could be developed to provide supportive prompts after 30 min of sitting, using Fitbit’s optimal threshold.

## Discrepancy between day level and short time scales

In line with previous research,<sup>21,49–55</sup> our study found that both Fitbits overestimated daily step count compared to the ActiGraph with MAPE of approximately 23.23% in adults and 27.50% in older adults. According to The Towards Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE) guidelines, a MAPE relative to a research criterion of  $\leq 10$ –15% is recommended to be a valid measure of step counts in the general population.<sup>38</sup> These discrepancies in step count between Fitbits and ActiGraph on a day level may be

**Table 4.** Sensitivity and specificity of Fitbits for stepping events for older adults (>65) as identified by the ActiGraph.

Duration (min)		60 steps/min		80 steps/min		100 steps/min		120 steps/min	
		Ionic	Inspire	Ionic	Inspire	Ionic	Inspire	Ionic	Inspire
1	Sensitivity (%)	98.22	98.16	97.74	98.31	95.36	95.80	51.73	51.73
	Specificity (%)	97.84	97.54	97.78	98.67	99.35	99.32	99.95	99.92
	Cohen's Kappa [95% CI] <sup>a</sup>	0.83 [0.82–0.84]	0.81 [0.80–0.82]	0.87 [0.86–0.88]	0.87 [0.86–0.88]	0.89 [0.88–0.90]	0.89 [0.88–0.90]	0.66 [0.63–0.69]	0.66 [0.63–0.69]
	Occurrence (n)	2928		2478		1855		866	
	Participants (n)	26		23		22		15	
5	Sensitivity (%)	99.41	99.26	98.87	99.01	96.24	97.68	39.97	42.90
	Specificity (%)	98.41	98.30	99.13	99.03	99.57	99.54	99.96	99.97
	Cohen's Kappa [95% CI]	0.86 [0.85–0.87]	0.85 [0.84–0.86]	0.90 [0.89–0.91]	0.89 [0.88–0.90]	0.91 [0.89–0.92]	0.90 [0.91–0.92]	0.55 [0.52–0.59]	0.59 [0.55–0.63]
	Occurrence (n)	2693		2115		1382		613	
	Participants (n)	24		20		17		8	
10	Sensitivity (%)	99.41	99.57	98.34	98.01	94.65	96.88	36.84	35.83
	Specificity (%)	98.61	98.48	99.20	99.09	99.65	99.60	99.97	99.98
	Cohen's Kappa [95% CI]	0.87 [0.87–0.88]	0.86 [0.87–0.88]	0.89 [0.88–0.90]	0.88 [0.87–0.89]	0.90 [0.89–0.91]	0.90 [0.89–0.91]	0.53 [0.48–0.57]	0.52 [0.47–0.56]
	Occurrence (n)	2544		1863		1121		494	
	Participants (n)	19		18		14		7	
20	Sensitivity (%)	99.20	99.38	98.13	98.07	92.17	96.14	33.83	35.93
	Specificity (%)	98.75	98.58	99.37	99.29	99.74	99.69	99.99	99.99
	Cohen's Kappa [95% CI]	0.88 [0.87–0.89]	0.87 [0.86–0.88]	0.90 [0.89–0.91]	0.89 [0.88–0.90]	0.89 [0.88–0.91]	0.90 [0.89–0.92]	0.50 [0.44–0.55]	0.52 [0.46–0.58]
	Occurrence (n)	2247		1552		856		334	
	Participants (n)	18		16		11		5	

<sup>a</sup>CI = 95% confidence interval.

attributed to different wear locations of the wearables: devices worn on the wrist may falsely recognise hand movements as steps.<sup>56</sup> Moreover, Tudor Locke et al. (2015) suggested that differences in mean steps/day may arise from differences in instrument sensitivity thresholds.<sup>43</sup>

In contrast to overestimation on a daily level, both Fitbits underestimated the step count on short time scales (5–

20 min) in older adults (but not in adults). It is a well-known fact that older adults walk with less force and shuffle with less arm movement; thus, their steps can easily be missed by Fitbit. This phenomenon counteracts the extra steps resulting from falsely recognised hand movements. On a daily level, walking forms only a minor fraction of a day; thus, the number of false steps accumulated throughout

the day commonly surpasses the number of missed steps during walking events, leading to overestimation. However, in our analysis of short time scales, we only included minutes with  $\geq 60$  steps representing true walking. Consequently, the participants did not accumulate too many false steps, and the missed steps manifested as an underestimation. While it seems like a plausible explanation, the design of our study cannot prove it, and future studies are needed to confirm this hypothesis. In any case, the difference between Fitbits and ActiGraph never exceeded three steps per minute, and the CCC between ActiGraph and both Fitbit devices ranged between 0.80 and 0.99 in adults and 0.85 and 0.90 in older adults, indicating poor to almost perfect agreement, which is consistent with previous research (on a 15-min epoch in adults).<sup>21</sup>

### Agreement of two Fitbit models

Fitbit Inspire 2 and Ionic showed a good agreement on a short time scale. However, the daily overestimation in steps between Fitbit and ActiGraph was consistently larger with Fitbit Inspire 2, indicating that Fitbit Inspire 2 is less accurate in measuring daily steps compared to Fitbit Ionic. A previous study also reported a significant difference in the daily step count (MAPE of 12%) between two generations of Fitbit devices (Fitbit Flex<sup>TM</sup> and Flex 2<sup>TM</sup>), finding an agreement of 0.95.<sup>57</sup> Despite these differences on a daily level, the validity and accuracy in detecting short bouts of stepping and SB on a short time scale were very similar, indicating that both devices can be used equivalently for event-based EMA or JITAIs. Given that Inspire 2 and Ionic well represent the wider Fitbit portfolio (Inspire 2 is a lower-end tracker, while Ionic is a high-end smartwatch) as well as older and newer product lines (Ionic was launched in 2017, and the product line has already been discontinued, while Inspire 2 was launched in 2020, and its revamped version Inspire 3 was released in late 2022), we can reasonably assume that the results can be generalised to other Fitbit models. Moreover, 100% accuracy may not be needed since the main goal of JITAIs and event-based EMA remains to increase PA or to examine within- and between-subject variation of determinants rather than performing diagnostic tests.

### Fitbit's potential for event-based EMA and JITAIs

Besides continuously sensing physical behaviour to automatically elicit event-based EMA prompts, other technological aspects are critical in determining whether Fitbit and similar commercial wearable sensors can be utilised in EMA research or JITAIs. Successfully connecting devices with EMA or JITAI platforms requires (1) real-time transmission of sensor data to the platform and (2) an application programming interface (API) provided by the device manufacturer to allow EMA platforms real-time access to

sensor data. Since Fitbit can be considered a suitable device for capturing short bouts of PA and SB, and Fitbit synchronises the device data with the Fitbit server, where it can be accessed through Fitbit API, Fitbit may be connectable to existing EMA platforms. Researchers can use commercially available platforms (e.g. Fitabase<sup>58</sup>) and research platforms (e.g. Health-React<sup>59</sup> and iCardia<sup>60</sup>) or develop their own data collection interface connected to the Fitbit server to gather large amounts of data using Fitbit's API.<sup>61</sup> However, frequent syncing is of utmost importance for adequately gathering real-time data and for preventing data loss. Therefore, continuously enabling Bluetooth and internet connection is required, which is one of the most common issues in studies using Fitbit and may negatively affect smartphone battery drain.<sup>62</sup>

### Strengths and limitations

This study has several strengths. First, this is the first study that explored the accuracy of Fitbits in detecting short bouts of stepping and SB with practical implications for event-based EMA studies and JITAIs. Validation of Fitbits on a short time scale enables monitoring PA in real time and therefore offers the possibility to serve as a wearable sensor to trigger event-based EMA questionnaires or to provide feedback at the right time and in the right context. Second, our study comprised a wide range of PA levels in our study sample of adults and older adults, which suggests a good representation of the general population. Third, data were collected in two different countries (i.e. Belgium and the Czech Republic), further improving the generalisability of the findings. Fourth, the inclusion of two Fitbit models allowed us to examine the inter-device agreement and their comparability in the accuracy of detecting short bouts of stepping and SB. Fifth, all minutes with a step rate below 60 were excluded to evaluate the validity of Fitbits in measuring stepping events. When comparing the Fitbit-recorded step count per minute with the ActivPAL classification for stepping minutes, all ActivPAL-recorded minutes that consisted entirely of stepping reached the threshold of 60 steps/min with Fitbit. This is consistent with previous research, showing that a walking pace equal to or higher than 60 steps/min may be considered sustained walking.<sup>40</sup>

This study also has some limitations. First, the hip-worn ActiGraph GT3X+ was used as a concurrent measure. Although this triaxial accelerometer is widely used in PA research and has been validated for step counting in both adults and older adults,<sup>63,64</sup> the golden standard for measuring steps remains direct observation.<sup>65</sup> However, this method is not feasible to use in free-living conditions. Second, there is a rapid emergence of new wearables on the market. This means that validating new wearables are still required, and researchers have to remain critical when using a new device. Nevertheless, the two different

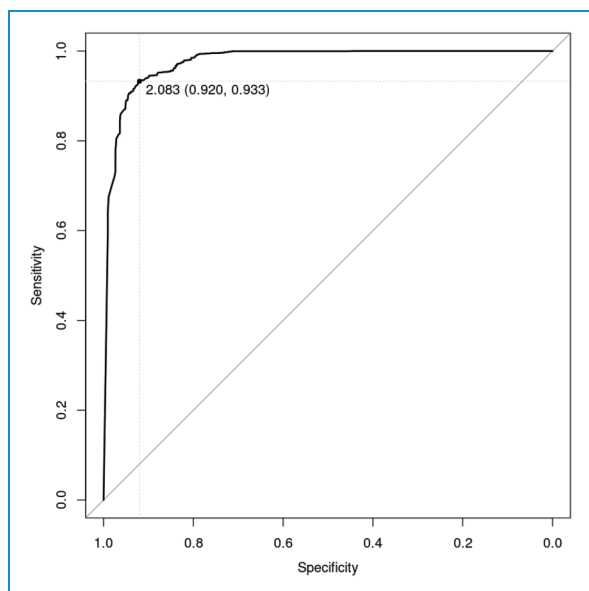
**Table 5.** Accuracy in detecting sedentary events as identified by ActivPAL.

Duration (min)		Adults		Older adults	
		Ionic	Inspire 2	Ionic	Inspire 2
1	Sensitivity (%)	93.94	93.26	95.54	94.82
	Specificity (%)	68.03	70.71	60.78	64.39
	Cohen's Kappa 95% CI	0.64 [0.63-0.64]	0.66 [0.65-0.66]	0.57 [0.56-0.58]	0.60 [0.59-0.60]
	Cut-off <sup>a</sup> (steps/min)	0.5	0.5	0.5	0.5
20	Sensitivity (%)	97.03	94.31	92.90	94.15
	Specificity (%)	89.55	93.16	93.01	91.52
	Cohen's Kappa 95% CI	0.87 [0.87-0.88]	0.87 [0.87-0.88]	0.86 [0.85-0.87]	0.86 [0.85-0.86]
	Cut-off <sup>a</sup> (steps/min)	2.58	1.73	1.08	1.98
30	Sensitivity (%)	97.31	95.34	97.67	93.27
	Specificity (%)	90.56	93.23	89.03	91.99
	Cohen's Kappa 95% CI	0.89 [0.88-0.89]	0.89 [0.88-0.90]	0.85 [0.84-0.86]	0.85 [0.84-0.86]
	Cut-off <sup>a</sup> (steps/min)	2.08	1.62	2.27	2.08
60	Sensitivity (%)	96.85	98.35	100	92.96
	Specificity (%)	96.88	96.10	88.71	89.36
	Cohen's Kappa 95% CI	0.94 [0.92-0.95]	0.94 [0.93-0.96]	0.81 [0.79-0.83]	0.77 [0.75-0.79]
	Cut-off <sup>a</sup> (steps/min)	1.07	1.54	2.11	2.6

<sup>a</sup>Cut-off identified by receiver operating characteristic (ROC) curves.

Fitbit models showed similar accuracies in detecting stepping and SB and had a good inter-device agreement on a short time scale, indicating that these results can be generalised to other Fitbit models. Third, this paper implies the use of steps/minute instead of specific intensity (e.g. MVPA) as a threshold for prompting questionnaires for event-based

EMA or future JITAs. We chose to avoid the PA intensity because previous research showed a low accuracy of Fitbit in measuring the PA intensity. On a 15-min level, Fitbit overestimated MVPA by 20%, and on a daily level, an underestimation of 30% to 153% was found.<sup>21,66,67</sup> In addition, no information is publicly available regarding Fitbit



**Figure 3.** ROC curve for detecting 30 min SB in older adults with Fitbit Inspire 2.

intensity cut-points, limiting the use of Fitbit-recorded intensity as a prompting trigger. Furthermore, the lower limits of MVPA have been consistently shown to be at around 100 steps/min in adults,<sup>68–70</sup> enabling the use of cadence as a proxy for PA intensity. Fourth, wrist-worn Fitbits are not able to differentiate between sitting and standing; thus, their accuracy might be different for people who stand frequently: quiet standing could potentially be identified as sitting by Fitbit, reducing the specificity. Fifth, the sampling frequency of the ActiGraph was set to 30 Hz, while Fitbit devices typically record at a default sampling frequency of 100 Hz. This may affect the comparability of data between devices and should be considered when interpreting the results. Finally, participants were recruited through convenience sampling, which may limit the generalisability of the findings to the wider population.

## Conclusions

Fitbit wearables appear to be valid for measuring steps on a short time scale (i.e. 1, 5, 10 and 20 min) compared to ActiGraph. This paper provides practical considerations for event-based EMA and JITAIs using Fitbit as a device to assess PA and SB in adults and older adults. We present an overview of the sensitivities and specificities for several step-rate thresholds to capture short bouts of stepping (1, 5, 10 and 20 min) and SB (1, 20, 30 and 60 min). Considering the sensitivity and specificity for a particular research purpose (e.g. 5 min of sustained walking or 30 min of sedentary time), Fitbits can be used as a wearable sensor to give real-time support. In addition, both Fitbit models showed a substantial agreement on short

time scales, indicating that these results may be generalised to other Fitbit models. Since Fitbits are relatively cheap devices, their use as a tool for triggering questionnaires or prompts in real time can enhance further development of event-based EMA methods and JITAIs for promoting PA and limiting SB.

**Conflict of Interest information:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Contributorship:** JD, EL, DVD and TV conceived the study. JD, EL and MJ were involved in protocol development, gaining ethical approval, patient recruitment and data collection. JD performed the statistical analyses and wrote the first draft of the manuscript. TV, DVD and SC verified the analyses and aided in interpreting the results. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

**Data availability:** The data sets and scripts generated during and/or analyzed during this study are available in the Open Science Framework repository (OSF) ([osf.io/jva8c](https://osf.io/jva8c)).<sup>71</sup>

**Ethical approval:** The study protocol was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Ghent University Hospital (BC-09448) and Charles University (No. 299/2021).

**Funding:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The current study has received financial support from the Czech Health Research Council, Ministry of Health of the Czech Republic (grant number NU21-09-00007) and from the Research Foundation — Flanders (FWO, project number 3G005520).

**Guarantor:** TV.

**ORCID ID:** Julie Delobelle  <https://orcid.org/0000-0003-3771-2848>

**Supplemental material:** Supplemental material for this article is available online.

## References

1. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012; 380: 219–229.
2. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *Br Med J* 2019; 366: 14570.

3. Schuch FB, Vancampfort D, Richards J, et al. Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. *J Psychiatr Res* 2016; 77: 42–51.
4. Rezende LFM, Sá TH, Mielke GI, et al. All-cause mortality attributable to sitting time: analysis of 54 countries worldwide. *Am J Prev Med* 2016; 51: 253–263.
5. Shrestha N, Grgic J, Wiesner G, et al. Effectiveness of interventions for reducing non-occupational sedentary behaviour in adults and older adults: a systematic review and meta-analysis. *Br J Sports Med* 2019; 53: 1206–1213.
6. Conn VS, Hafdahl AR and Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. *Am J Public Health* 2011; 101: 751–758.
7. Müller-Riemenschneider F, Reinhold T, Nocon M, et al. Long-term effectiveness of interventions promoting physical activity: a systematic review. *Prev Med* 2008; 47: 354–368.
8. Dishman RK and Buckworth J. Increasing physical activity: a quantitative synthesis. 2007.
9. Spence JC and Lee RE. Toward a comprehensive model of physical activity. *Psychol Sport Exerc* 2003; 4: 7–24.
10. Conroy DE, Maher JP, Elavsky S, et al. Sedentary behavior as a daily process regulated by habits and intentions. *Health Psychol* 2013; 32: 1149–1157.
11. Maes I, Mertens L, Poppe L, et al. The variability of emotions, physical complaints, intention, and self-efficacy: an ecological momentary assessment study in older adults. *PeerJ* 2022; 10: e13234.
12. Dunton GF. Ecological momentary assessment in physical activity research. *Exerc Sport Sci Rev* 2017; 45: 48–54.
13. Nahum-Shani I, Smith S, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAI): an organizing framework for ongoing health behavior support: (Technical Report No. 14-126). University Park, PA: The Methodology Center, Penn State, 2014.
14. Hardeman W, Houghton J, Lane K, et al. A systematic review of Just-in-Time Adaptive Interventions (JITAI) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity* 2019; 16: 31.
15. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAI) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018; 52: 446–462.
16. Stone AA and Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med* 1994; 16: 199–202.
17. Schüz N, Cianchi J, Shiffman S, et al. Novel technologies to study smoking behavior: current developments in ecological momentary assessment. *Curr Addict Rep* 2015; 2: 8–14.
18. Kaplan RM and Stone AA. Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention. *Annu Rev Psychol* 2013; 64: 471–498.
19. Shiffman S, Stone AA and Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008; 4: 1–32.
20. Ringeval M, Wagner G, Denford J, et al. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. *J Med Internet Res* 2020; 22: e23954–e23954.
21. Degroote L, De Bourdeaudhuij I, Verloigne M, et al. The accuracy of smart devices for measuring physical activity in daily life: validation study. *JMIR Mhealth Uhealth* 2018; 6: e10972.
22. Dunton GF, Dzibur E and Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. *J Med Internet Res* 2016; 18: e106.
23. Giurgiu M, Timm I, Becker M, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: systematic review. *JMIR Mhealth Uhealth* 2022; 10: e36377.
24. Keadle SK, Lyden KA, Strath SJ, et al. A framework to evaluate devices that assess physical behavior. *Exerc Sport Sci Rev* 2019; 47: 206–214.
25. Evenson KR, Goto MM and Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity* 2015; 12: 59.
26. Feehan LM, Geldman J, Sayre EC, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth Uhealth* 2018; 6: e10527.
27. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med* 2013; 34: 975–982.
28. Storti KL, Pettee KK, Brach JS, et al. Gait speed and step-count monitor accuracy in community-dwelling older adults. *Med Sci Sports Exerc* 2008; 40: 59–64. Article; Proceedings Paper.
29. Kim Y, Barry VW and Kang M. Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. *Meas Phys Educ Exerc Sci* 2015; 19: 125–137.
30. Kozey-Keadle S, Libertine A, Lyden K, et al. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exercise* 2011; 43: 1561–1567.
31. Godfrey A, Culhane KM and Lyons GM. Comparison of the performance of the activPAL professional physical activity logger to a discrete accelerometer-based activity monitor. *Med Eng Phys* 2007; 29: 930–934.
32. Hart TL, Ainsworth BE and Tudor-Locke C. Objective and subjective measures of sedentary behavior and physical activity. *Med Sci Sports Exercise* 2011; 43: 449–456.
33. Ryan CG, Grant PM, Tigbe WW, et al. The validity and reliability of a novel activity monitor as a measure of walking. *Br J Sports Med* 2006; 40: 779–784.
34. Johnston W, Judice PB, García PM, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med* 2021; 55: 780–793.
35. Github. teramonagi fitbitr.
36. Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008; 40: 181–188.
37. O'Brien MW, Pellerine LP, Shivgulam ME, et al. Disagreements in physical activity monitor validation study guidelines create challenges in conducting validity studies. *Front Digit Health* 2022; 4: 1063324.
38. Johnston W, Judice PB, Molina García P, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med* 2021; 55: 780–793.

39. GB M. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062. National Institute of Water & Atmospheric Research: Hamilton, New Zealand; 2005.
40. Dall PM, McCrorie PR, Granat MH, et al. Step accumulation per minute epoch is not the same as cadence for free-living adults. *Med Sci Sports Exerc* 2013; 45: 1995–2001.
41. Watson PF and Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology* 2010; 73: 1167–1179.
42. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32–35.
43. Tudor-Locke C, Barreira TV and Schuna Jr JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sci Sports Exerc* 2015; 47: 839–842.
44. Romkes J and Bracht-Schweizer K. The effects of walking speed on upper body kinematics during gait in healthy subjects. *Gait Posture* 2017; 54: 304–310.
45. Degroote L, DeSmet A, De Bourdeaudhuij I, et al. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity* 2020; 17: 35.
46. Maher J, Rebar A and Dunton G. Ecological momentary assessment is a feasible and valid methodological tool to measure older adults' physical activity and sedentary behavior. *Front Psychol* 2018; 9: 1485.
47. Rutten GM, Savelberg HH, Biddle SJ, et al. Interrupting long periods of sitting: good STUFF. *Int J Behav Nutr Phys Act* 2013; 10: 1.
48. WHO guidelines on physical activity and sedentary behaviour: Web Annex. *Evidence profiles*. Geneva: World Health Organization, 2020. Licence: CC BY-NC-SA 3.0 IGO.
49. Schneider M and Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Res Notes* 2016; 9: 48.
50. Tully MA, McBride C, Heron L, et al. The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Res Notes* 2014; 7: 52.
51. Dominick GM, Winfree KN, Pohlig RT, et al. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. *JMIR Mhealth Uhealth* 2016; 4: e110.
52. Bunn JA, Navalta JW, Fountaine CJ, et al. Current state of commercial wearable technology in physical activity monitoring 2015-2017. *Int J Exerc Sci* 2018; 11: 503–515.
53. Matlary RED, Holme PA, Glosli H, et al. Comparison of free-living physical activity measurements between ActiGraph GT3X-BT and Fitbit charge 3 in young people with haemophilia. *Haemophilia* 2022; 28: e172–e180.
54. Chu AHY, Ng SHX, Paknezhad M, et al. Comparison of wrist-worn Fitbit flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLOS ONE* 2017; 12: e0172535.
55. Germini F, Noronha N, Borg Debono V, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. *J Med Internet Res* 2022; 24: e30791.
56. Nelson R, Hasanaj K, Connolly G, et al. Comparison of wrist- and hip-worn activity monitors when meeting step guidelines. *Prev Chronic Dis* 2022; 19: E18.
57. Jones DM, Hart HF, Crossley KM, et al. What is the agreement between two generations of commercial accelerometer in a free-living environment for young to middle-aged adults? *Journal for the Measurement of Physical Behaviour* 2019; 2: 49–57.
58. Fitabase. 2023.
59. Vetrovsky T, Kral N, Pfeiferova M, et al. mHealth intervention delivered in general practice to increase physical activity and reduce sedentary behaviour of patients with prediabetes and type 2 diabetes (ENERGISED): rationale and study protocol for a pragmatic randomised controlled trial. *BMC Public Health* 2023; 23: 13.
60. Kitsiou S, Thomas M, Marai GE, et al. Development of an innovative mHealth platform for remote physical activity monitoring and health coaching of cardiac rehabilitation patients. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) 16-19 Feb. 2017 2017, pp.133–136.
61. Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *Br J Sports Med* 2021; 55: 422–432.
62. Balbim GM, Marques IG, Marquez DX, et al. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. *JMIR Mhealth Uhealth* 2021; 9: e25289.
63. Ngueleu A-M, Barthod C, Best KL, et al. Criterion validity of ActiGraph monitoring devices for step counting and distance measurement in adults and older adults: a systematic review. *J Neuroeng Rehabil* 2022; 19: 12.
64. Lee JA, Williams SM, Brown DD, et al. Concurrent validation of the ActiGraph GT3X+, polar active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *J Sports Sci* 2015; 33: 991–1000.
65. Sirard J and Pate R. Physical activity assessment in children and adolescents. *Sports Medicine (Auckland, NZ)* 2001; 31: 439–454.
66. Ferguson T, Rowlands AV, Olds T, et al. 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: 42.
67. Reid RER, Insogna JA, Carver TE, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *J Sci Med Sport* 2017; 20: 578–582.
68. Slaght J, Sénéchal M, Hrubeniuk TJ, et al. Walking cadence to exercise at moderate intensity for adults: a systematic review. *J Sports Med (Hindawi Publ Corp)* 2017; 2017: 4641203.
69. Rowe D, Welk G, Heil D, et al. Stride rate recommendations for moderate-intensity walking. *Med Sci Sports Exerc* 2011; 43: 312–318.
70. Tudor-Locke C, Sisson SB, Collova T, et al. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Canadian Journal of Applied Physiology* 2005; 30: 666–676.
71. Delobelle J. *Data [Internet]*. Ghent, Belgium: OSF, 2023. Available from: [osf.io/jva8c](https://osf.io/jva8c).