The validity of activity trackers is affected by walking speed: the criterion validity of Garmin Vivosmart® HR and StepWatch™ 3 for measuring steps at various walking speeds under controlled conditions

Frederik Rose Svarre[1,2], Mads Møller Jensen[1,3], Josephine Nielsen[1] and Morten Villumsen[4,5]

[1] Department of Physiotherapy, University College of Northern Denmark, Aalborg, Denmark
[2] Department of Health and Movement, Jammerbugt Municipality, Pandrup, Denmark
[3] Department of Physiotherapy, Aalborg University Hospital, Hobro, Denmark
[4] Department of Elderly and Health, Section of Training and Activity, Aalborg Municipality, Aalborg, Denmark
[5] Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

ABSTRACT

Introduction: The use of activity trackers has increased both among private consumers and in healthcare. It is therefore relevant to consider whether a consumer-graded activity tracker is comparable to or may substitute a research-graded activity tracker, which could further increase the use of activity trackers in healthcare and rehabilitation. Such use will require knowledge of their accuracy as the clinical implications may be significant. Studies have indicated that activity trackers are not sufficiently accurate, especially at lower walking speeds. The present study seeks to inform decision makers and healthcare personnel considering implementing physical activity trackers in clinical practice. This study investigates the criterion validity of the consumer-graded Garmin Vivosmart® HR and the research-graded StepWatch™ 3 compared with manual step count (gold standard) at different walking speeds under controlled conditions.

Methods: Thirty participants, wearing Garmin Vivosmart® HR at the wrist and StepWatch™ 3 at the ankle, completed six trials on a treadmill at different walking speeds: 1.6 km/h, 2.4 km/h, 3.2 km/h, 4.0 km/h, 4.8 km/h, and 5.6 km/h. The participants were video recorded, and steps were registered by manual step count. Medians and inter-quartile ranges (IQR) were calculated for steps and differences in steps between manually counted steps and the two devices. In order to assess the clinical relevance of the tested devices, the mean absolute percentage error (MAPE) was determined at each speed. A MAPE ≤3% was considered to be clinically irrelevant. Furthermore, differences between manually counted steps and steps recorded by the two devices were presented in Bland–Altman style plots.

Results: The median of differences in steps between Garmin Vivosmart® HR and manual step count ranged from −49.5 (IQR = 101) at 1.6 km/h to −1 (IQR = 4) at 4.0 km/h. The median of differences in steps between StepWatch™ 3 and manual step count were 4 (IQR = 14) at 1.6 km/h and 0 (IQR = 1) at all other walking speeds. The results of the MAPE showed that differences in steps counted by Garmin

Vivosmart® HR were clinically irrelevant at walking speeds 3.2–4.8 km/h (MAPE: 0.61–1.27%) as the values were below 3%. Differences in steps counted by StepWatch™ 3 were clinically irrelevant at walking speeds 2.4–5.6 km/h (MAPE: 0.08–0.35%).

Conclusion: Garmin Vivosmart® HR tended to undercount steps compared with the manual step count, and StepWatch™ 3 slightly overcounted steps compared with the manual step count. Both the consumer-graded activity tracker (Garmin Vivosmart® HR) and the research-graded (StepWatch™ 3) are valid in detecting steps at selected walking speeds in healthy adults under controlled conditions. However, both activity trackers miscount steps at slow walking speeds, and the consumer graded activity tracker also miscounts steps at fast walking speeds.

INTRODUCTION

Physical activity, such as walking, is beneficial to the individuals' health, and physical activity is known to reduce the risk of musculoskeletal disorders, chronic diseases, and death (*Takacs et al., 2014*; *Kumahara et al., 2015*; *Riel et al., 2016*). Common recommendations for healthy adults regarding physical activity are either 30 min of moderate activity or walking 10,000 steps a day (*Haskell et al., 2007*; *Tudor-Locke et al., 2011b*). Consumer-graded off-the-shelf activity trackers are widely available for private use, and wearable devices were the top fitness trend worldwide in both 2016, 2017, 2019, and 2020 (*Thompson, 2015*, *2016*, *2018*, *2019*). Activity trackers are marketed as objective tools for measuring steps and activity, allowing individuals to monitor daily activity and, for example, get reminders to increase their activity to reach recommended or personalized goals (*Takacs et al., 2014*; *Evenson, Goto & Furberg, 2015*; *Kooiman et al., 2015*; *Huang et al., 2016*). Activity trackers seem to play a prominent role not only in objectifying individuals' physical activity patterns but also in demonstrating and monitoring changes in activity behavior (*Fokkema et al., 2017*). As feedback on activity level and personalized goals might be very important to users, the accuracy of activity trackers is an important element to consider (*Burton et al., 2018*; *Gaz et al., 2018*).

Recent years have seen a growing use of activity trackers among private consumers and in healthcare (*Fokkema et al., 2017*) where objective measures of activity can be used for self-management purposes, prediction of fall risk, and to measure the effect of rehabilitation programs, etc. (*Chigateri et al., 2018*). In healthcare, the accuracy of activity trackers must be known as lack of accuracy may have significant clinical implications when using activity trackers for various tasks such as monitoring and evaluating patient progress, supporting the choice of or adjusting interventions, or for machine learning purposes (*Ferguson et al., 2015*; *Treacy et al., 2017*). Activity trackers are typically divided into two groups; consumer-graded activity trackers and research-graded activity trackers. One might assume that the latter are better at detecting physical activity, such as steps,

than the often cheaper consumer-graded activity trackers (*Tudor-Locke et al., 2011a*; *Ferguson et al., 2015*). However, due to the expenses and user requirements of research-graded activity trackers, it is relevant to consider whether a consumer-graded activity tracker is comparable to or may substitute a research-graded activity tracker, which could increase the use of activity trackers in healthcare and rehabilitation.

Studies have indicated that activity trackers are not sufficiently accurate, especially at lower walking speeds (*Crouter et al., 2003*; *Ryan et al., 2006*; *Steeves et al., 2011*; *Van Remoortel et al., 2012*; *Fortune et al., 2014*; *Lee et al., 2015*; *Huang et al., 2016*; *Fokkema et al., 2017*; *Tedesco et al., 2019*). When using activity trackers in healthcare, the target populations are likely to have different walking speeds with intermittent decreases in walking (e.g., in rehabilitation after ligament tears) or continued slow walking (e.g., older or disabled people). This emphasizes the need to be certain about the accuracy of the activity trackers when using the data in care and treatment of patients (*Fokkema et al., 2017*). Thus, lack of accuracy, for example, in the form of overestimation or underestimation of the number of steps, may interfere with healthcare interventions and lead to untoward outcomes. Bearing this in mind, the accuracy of activity trackers must be validated (*Lee et al., 2015*; *Huang et al., 2016*). Previous validation studies have proposed and used different walking speeds. Some have used walking speeds between 2.4 km/h and 6.4 km/h, with intervals of 0.8 km/h (*Edbrooke et al., 2012*; *De Cocker et al., 2012*; *Lee et al., 2015*; *Gaz et al., 2018*). However, older or disabled people can have even lower walking speed than 2.4 km/h (*Graham et al., 2010*). The aim of the present study was to determine differences in step count of an off-the-shelf consumer-graded activity tracker and a research-graded activity tracker to investigate the criterion validity. This study investigates the criterion validity of the wrist-worn, consumer-graded Garmin Vivosmart® HR and the ankle-worn, research-graded StepWatch™ 3 compared with manual step count (gold standard) at different walking speeds under controlled conditions. By adding to the knowledge base on the criterion validity under controlled conditions, this study seeks to inform and support decision makers and healthcare personnel when considering implementing physical activity trackers in clinical practice.

METHODS

This criterion validity study complies with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (*Kottner et al., 2011*) and follows the taxonomy, terminology, and definitions from the "COnsensus-based Standards for the selection of health status Measurement INstruments" (the COSMIN study) (*Mokkink et al., 2010a*, *2010b*).

The choice of sample size, inclusion and exclusion criteria, and treadmill speeds was guided by the study aim and informed by previous study designs; the latter to facilitate comparison with other research in this field.

We find it important to investigate the criterion validity in healthy adults under controlled conditions prior to commencing a study in free-living conditions or in a non-healthy population.

Table 1 Demographical data.	
Gender (N, men/women)	12/18
Age (years, mean (SD))	26.6 (±6.2)
Height (cm, mean (SD))	173.9 (±10.7)
Weight (kg, mean (SD))	73.8 (±12.4)

Ethics statement

The North Denmark Region Committee on Health Research Ethics stated that according to ethics committee law number 593 of 14/6/2011 § 2, 1-3 and § 14, 1 (Confirmation date 02.16.2016), ethical approval was not required for this study. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Sample size and method

The sample size was set to 30 participants, which was in accordance with equivalent laboratory condition validation studies (*Takacs et al., 2014*; *Kooiman et al., 2015*; *Riel et al., 2016*). Two observers conducted data collection, manual step count, data processing, and statistical analysis. The observers were blinded to the results until all raw data had been processed and were ready for statistical analysis.

Participants

Thirty healthy adults (*n* = 18 females, *n* = 12 males), Table 1, were recruited through the University College of Northern Denmark by using a mailing list with contact addresses of students and staff. Verbal and written information on the study procedures and aim was provided, and written informed consent was obtained from all participants prior to commencing the study. All participation was voluntary. Data collection was conducted in April 2016.

Only healthy participants were included. The inclusion and exclusion criteria were in line with those of previous studies (*Hendrick et al., 2010*; *Steeves et al., 2011*; *Bergman et al., 2012*; *Fortune et al., 2014*, *2015*; *Takacs et al., 2014*; *Johnson et al., 2015*; *Stansfield, Hajarnis & Sudarshan, 2015*; *Hickey et al., 2015*). To ensure physical activity participation clearance participants were eligible for inclusion if they were: (i) aged 18–64 years (self-reported), (ii) able to read and comprehend Danish and English (self-reported), and (iii) responded "No" to all questions (7 out of 7) in the Physical Activity Readiness Questionnaire (PAR-Q) (*Shephard, 1988*; *Steeves et al., 2011*; *Johnson et al., 2015*). Moreover, (iv) for safety reasons, participants should be able to walk independently >40 min (self-reported) to ensure that they were able to complete the treadmill walking exercise without risking adverse events (e.g., injury). Criteria for exclusion were: (i) neurological diseases (self-reported), (ii) cognitive problems (self-reported), (iii) any musculoskeletal injuries and/or operation (self-reported) that might have affected walking six weeks prior to participation, and (iv) vascular issues (self-reported) since these may influence walking ability (*Takacs et al., 2014*; *Stansfield, Hajarnis & Sudarshan, 2015*; *Fortune et al., 2015*).

Garmin Vivosmart® HR

The Garmin Vivosmart® HR (Garmin, Olathe, Kansas, USA: 149,99 USD (*Garmin, 2016a*) is a wrist-worn, accelerometer-based off-the-shelf activity tracker monitoring step count, distance, calories, heart rate, number of floors climbed and intensity minutes. The Garmin Vivosmart® HR has a build-in display for direct operation and reading of data. Data storage and analysis from the Garmin Vivosmart® HR can be assessed by the user through a Garmin Connect™ account enabling optional settings (e.g., Health & Fitness, Units). (*Garmin, 2016b*). To configure the Garmin Vivosmart® HR through Garmin Express™ v. 4.1.19.0, a Garmin Connect™ account was created with the Health & Fitness profile enabled. The Garmin Connect™ account was solely used to login to the Garmin Express™ v. 4.1.19.0 software to configure the activity tracker for each participant individually (*Garmin, 2016b*, *2016c*). The Garmin Vivosmart® HR was configured individually for each participant by entering his or her date of birth, gender, weight, and height. Metric values were used during this configuration process, time was presented in 24-h format, and the lowest activity level (standard value) was chosen for all participants. The standard option for the wearing side for the activity tracker was set to "left" in the Garmin Express™ v. 4.1.19.0 software. The Garmin Vivosmart® HR was placed proximally from the caput ulnae sinister on each participant, as recommended by the manufacturer (Fig. 1) (*Garmin, 2016c*).

StepWatch™ 3

The StepWatch™ 3 (Modus Health, LLC, Edmonds, WA, USA: 525 USD (Accelerometer), 1470 USD (Docking station and software) (*Tudor-Locke et al., 2011a*)) is an ankle-worn, accelerometer-based, research-graded activity tracker monitoring step count and activity levels based on step rate. A docking station and software are needed to configure the StepWatch™ 3 and access the recorded data (*Tudor-Locke et al., 2011a*; *Modus Health, 2014*). The StepWatch™ 3 was configured individually for each participant by entering gender, height, weight, and age. Furthermore, the recording interval was defined to 15 s and Client ID and Trial ID were entered for subsequent data analysis. The two StepWatch™ 3 monitors (one on each leg) were mounted according to the manufacturer's recommendation just proximal from the malleolus lateralis of the ankle (Fig. 2) (*Modus Health, 2014*).

Video recording and manual step count

Manual step count is regarded the gold standard for step counting during treadmill walking in laboratory conditions (*Tudor-Locke et al., 2002*; *Fortune et al., 2014*, *2015*; *Lee et al., 2015*; *Stansfield, Hajarnis & Sudarshan, 2015*; *Sellers et al., 2016*). Thus, manual step count was used during walking on a zebris FDM-T treadmill (zebris Medical GmbH, Isny, Germany). In this study, a GoPro Hero 4 Black Edition (GoPro, Inc., San Mateo, CA, USA) was used to record treadmill walking. The camera was placed so that the participant's sagittal plane was recorded.

Figure 1 Illustration of the placement of Garmin Vivosmart® HR.
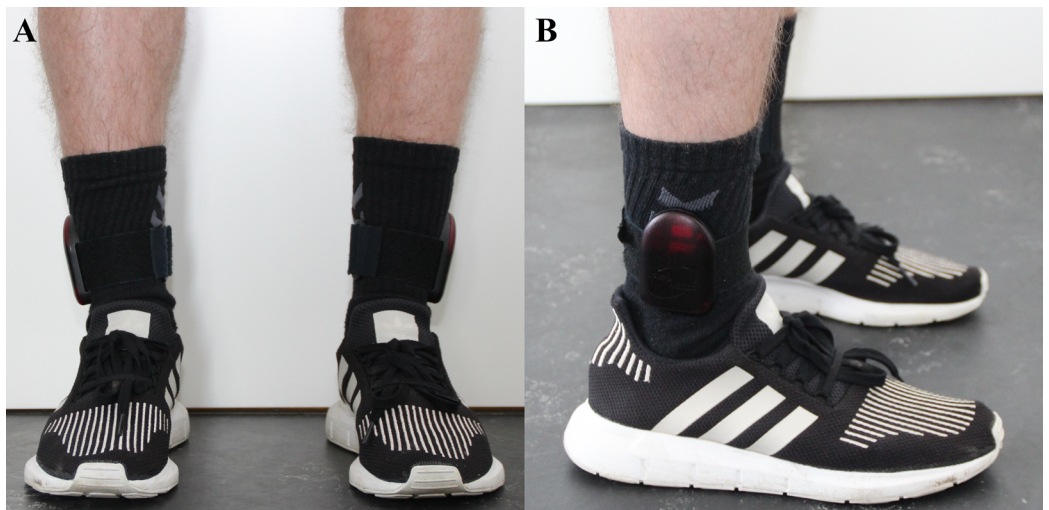Full-size ⬜ DOI: 10.7717/peerj.9381/fig-1

Figure 2 Illustration of the placement of StepWatch™ 3 monitors. (A) Frontal view, (B) Sagittal view.
Full-size ⬜ DOI: 10.7717/peerj.9381/fig-2

In this study, a step was defined in accordance with Hickey and colleagues as "when the foot was completely raised off and subsequently lowered to the ground" (*Hickey et al., 2015*). Using this definition, the two observers individually viewed the recordings using

VLC version 2.2.4 (VideoLAN, Paris, France) while using a hand tally (Rucanor BSI ISO 09001 Certified No. FM40047) ensuring blinding and double validation of the manual step counting based on the approach in similar studies (*Rosenkranz, Rosenkranz & Weber, 2011*; *Aminian & Hinckson, 2012*; *Takacs et al., 2014*; *Riel et al., 2016*). In cases of discrepancies between the two observers' step counts, a new counting was performed to ensure valid counts as done by *Takacs et al. (2014)*.

Procedure of the treadmill test

In order to examine the criterion validity of Garmin Vivosmart® HR and StepWatch™ 3 in step counting under controlled conditions, the participants were asked to perform six trials with a total sum of 40 minutes of treadmill walking at various speeds. Thus, to capture both very low, low, moderate, and high walking speeds, paces of 1.6 km/h, 2.4 km/h, 3.2 km/h, 4.0 km/h, 4.8 km/h, and 5.6 km/h were chosen, based on the approaches of previous studies (*Edbrooke et al., 2012*; *De Cocker et al., 2012*; *Lee et al., 2015*; *Gaz et al., 2018*). However, none of these studies included walking speeds below 2.4 km/h. We therefore found it particularly relevant to include the 1.6 km/h walking speed in this study as the use of activity trackers in healthcare may include target populations with intermittent decreases in walking speed (e.g., rehabilitation after ligament tears) or continued slow walking speeds (e.g., older or disabled people). In line with previous studies, each participant familiarized him or herself with the equipment before the actual test (*Ryan et al., 2006*; *Giannakidou et al., 2012*), walking the six different speeds for 2 min each to gain experience with the acceleration and deceleration of the treadmill.

As suggested in previous studies, a test period of 5 min per speed was used (*Le Masurier & Tudor-Locke, 2003*; *Crouter et al., 2003*; *Ryan et al., 2006*; *Giannakidou et al., 2012*; *De Cocker et al., 2012*; *Takacs et al., 2014*). An interval time break of 90 s was used to allow time to record data from the Garmin Vivosmart® HR and make a break in the recording of the StepWatch™ 3. The procedure for the interval time break was based on previous treadmill walking studies (*Bergman et al., 2012*; *Lee et al., 2015*). Increases and decreases in speed in each trial were programed to 0.91 m/s and were included in the 5-min test period. For safety reasons, verbal countdown from the observer was given 10, 5, and 3 s prior to the starting and stopping event. The treadmill was set to an elevation of 0 degrees verified by a spirit level, in line with similar studies (*Le Masurier & Tudor-Locke, 2003*; *Steeves et al., 2011*). As recommended by the manufacturer, service and calibration of the treadmill's pressure sensors and speedometer were performed prior to this study by an employe at *zebris Medical GmbH (2015)*. The tests were carried out in the Movement Laboratory, Department of Physiotherapy, University College of Northern Denmark, Aalborg, Denmark.

Processing data

Garmin Vivosmart® HR: Start and stop values of step count recorded by the activity tracker were collected manually before and after every trial and transferred into a

Microsoft Office Excel Datasheet. The actual number of steps recorded by the activity tracker was calculated by subtracting the start value from the stop value.

StepWatch™ 3: When the participant had completed all trials, data were downloaded for both StepWatch™ 3 activity trackers with software provided by the manufacturer—StepWatch™ v.3.4. Data were exported with the software StepWatch™ v.3.4 into a Microsoft Office Excel Datasheet, where the sum of recorded steps by the two StepWatch™ 3 activity trackers was identified for each trial.

Video data from the GoPro Hero 4 Black Edition were transferred using the associated application/software GoPro Desktop v1.3.0.2371, available from the manufacturer's web page (*GoPro, 2016*). Data from each walking speed were registered using a visual speed indicator and saved as separate files for subsequent identification. The video recordings were used as the underlying technical basis for manual step count.

Statistical analysis

All statistical analyses were conducted in *R* version 3.6.1 and *RStudio* version 1.2.1335. The criterion validity of Garmin Vivosmart® HR and StepWatch™ 3 was established by a series of statistical tests applying a significance level of $p < 0.05$.

The Shapiro–Wilk test was used to determine the distribution of normality of the differences between Garmin Vivosmart® HR and manual step count, and between StepWatch™ 3 and manual step count. The Shapiro–Wilk test showed events of non-normality ($p < 0.05$); therefore, non-parametric tests were applied for further statistical analysis. Medians and inter-quartile ranges (IQR) were calculated by the default settings in *R* for steps and differences in steps between manually counted steps and the two devices.

In order to assess the clinical relevance of the tested devices, the mean absolute percentage error (MAPE) was determined at each walking speed for Garmin Vivosmart® HR and StepWatch™ 3. The MAPE was calculated as $\text{MAPE} = \frac{|\text{Activity tracker} - \text{Manual Count}|}{\text{Manual Count}} \times 100\%$. Earlier studies have described a MAPE ≤3% to be clinically irrelevant (*Holbrook, Barreira & Kang, 2009*; *Colley et al., 2013*; *Johnson et al., 2015*; *Liu et al., 2015*; *Riel et al., 2016*; *Bunn et al., 2018*). The differences between manually counted steps and steps recorded by the two devices were presented in Bland Altman style plots along with the median, the 2.5th, and the 97.5th quartile due to the non-normally distributed dataset (*Gialamas et al., 2010*; *Riel et al., 2016*). All data were plotted, and no potential outliers were removed.

RESULTS

Garmin Vivosmart® HR vs. manual step count

The medians of differences in steps between Garmin Vivosmart® HR and manual step count were −49.5 (IQR = 101) at 1.6 km/h, −7.5 (IQR = 14) at 2.4 km/h, −2.5 (IQR = 6) at 3.2 km/h, −1 (IQR = 4) at 4.0 km/h, −1 (IQR = 5) at 5.8 km/h, and −4 (IQR = 7) at 5.6 km/h (Table 2). These are plotted in the Bland–Altman style plot (Fig. 3).

The results of the MAPE showed that the steps counted by Garmin Vivosmart® HR had clinically relevant deviations at 1.6 km/h, 2.4 km/h, and 5.6 km/h (MAPE: 3.49–26.35%) compared with manual step count (Table 2). The differences in step count

Table 2 The median of counted steps by device, and the median of differences in steps by device compared to manual step count, and the mean absolute percentage error by device compared to manual step count.

Device	Walking speed (km/h)	Median (steps)	Median of difference (steps)	Mean absolute percentage error (%) ± SD
Manual step count	1.6	337.5 (IQR = 33)	–	–
	2.4	410 (IQR = 29)	–	–
	3.2	473 (IQR = 38)	–	–
	4.0	529.5 (IQR = 40)	–	–
	4.8	574 (IQR = 46)	–	–
	5.6	608 (IQR = 49)	–	–
Garmin Vivosmart® HR	1.6	292.5 (IQR = 106)	−49.5 (IQR = 101)	26.35 (±23.63)
	2.4	398 (IQR = 34)	−7.5 (IQR = 14)	3.49 (±4.10)
	3.2	465 (IQR = 36)	−2.5 (IQR = 6)	1.27 (±1.67)
	4.0	527 (IQR = 37)	−1 (IQR = 4)	0.61 (±0.61)
	4.8	570 (IQR = 45)	−1 (IQR = 5)	0.61 (±0.68)
	5.6	577 (IQR = 53)	−4 (IQR = 47)	6.45 (±11.46)
StepWatch™ 3	1.6	345 (IQR = 22)	4 (IQR = 14)	3.60 (±6.03)
	2.4	412 (IQR = 29)	0 (IQR = 1)	0.35 (±1.00)
	3.2	473.5 (IQR = 36)	0 (IQR = 1)	0.09 (±0.12)
	4.0	529.5 (IQR = 40)	0 (IQR = 1)	0.08 (±0.09)
	4.8	574 (IQR = 47)	0 (IQR = 1)	0.08 (±0.10)
	5.6	608.5 (IQR = 49)	0 (IQR = 1)	0.09 (±0.12)

for the device were interpreted as being clinically irrelevant as the values were below 3% at walking speeds of 3.2–4.8 km/h (MAPE: 0.61–1.27%).

StepWatch™ 3 vs. manual step count

The medians of differences in steps between StepWatch™ 3 and manual step count were 4 (IQR = 14) at 1.6 km/h and 0 (IQR = 1) at 2.4 km/h, 3.2 km/h, 4.0 km/h, 4.8 km/h, and 5.6 km/h (Table 2). These are plotted in the Bland–Altman style plot (Fig. 4).

The results of the MAPE showed that the steps counted by StepWatch™ 3 had a clinically relevant deviation at 1.6 km/h (MAPE: 3.60%) compared with manual step count (Table 2). The differences in step count for the device were interpreted clinically irrelevant as the values were below 3% at walking speeds of 2.4–5.6 km/h (MAPE: 0.08–0.35%).

DISCUSSION

This study aimed to investigate the criterion validity of Garmin Vivosmart® HR and StepWatch™ 3 at different walking speeds under controlled conditions. The results showed a tendency towards Garmin Vivosmart® HR undercounting steps compared with manual step count, and StepWatch™ 3 slightly overcounted steps compared with manual step count.
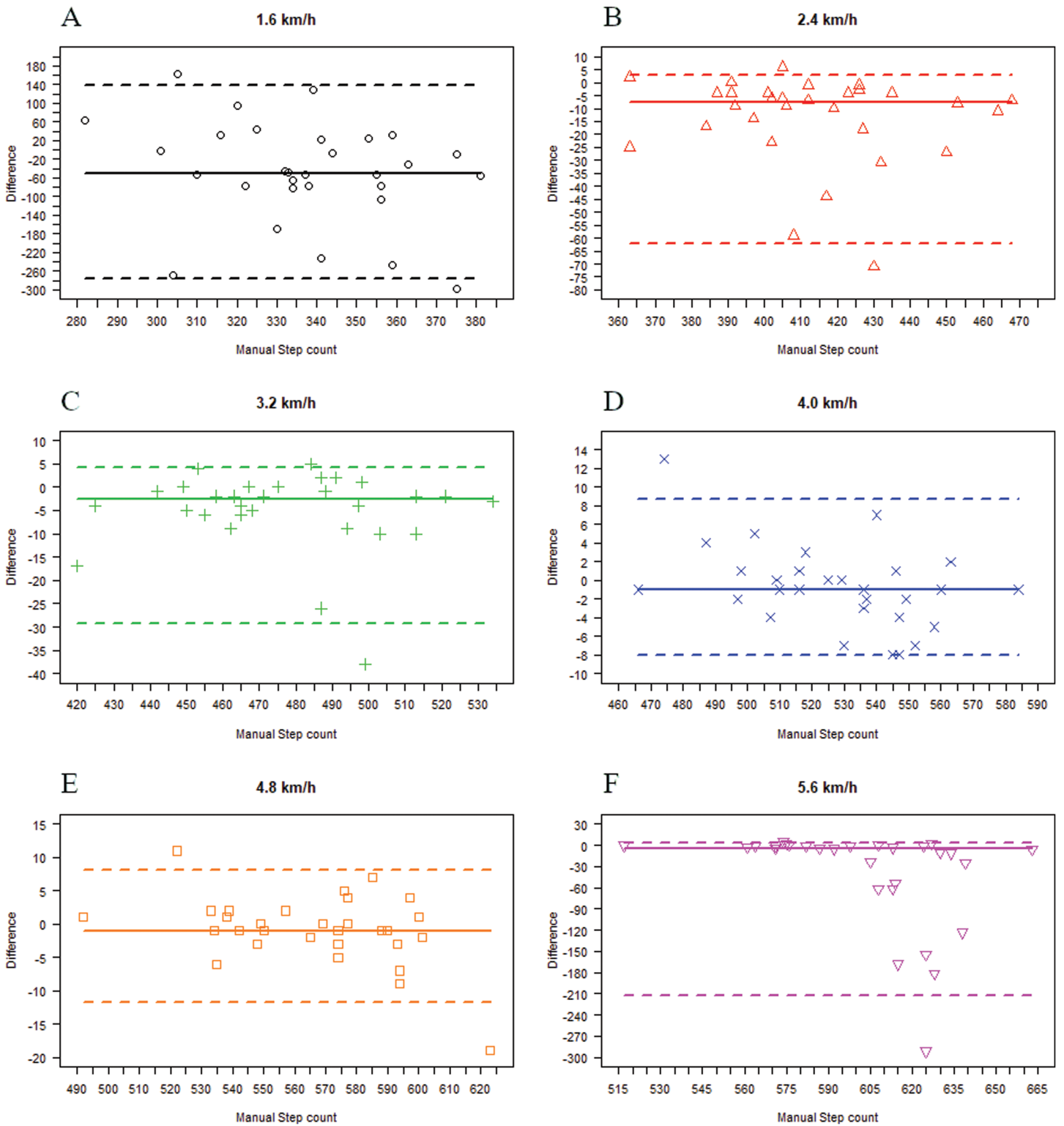
Figure 3 Bland Altman style plot illustrating the differences between Garmin Vivosmart® HR and manual step count at walking speeds from 1.6 km/h to 5.6 km/h. (A) 1.6 km/h, (B) 2.4 km/h, (C) 3.2 km/h, (D) 4.0 km/h, (E) 4.8 km/h, (F) 5.6 km/h.
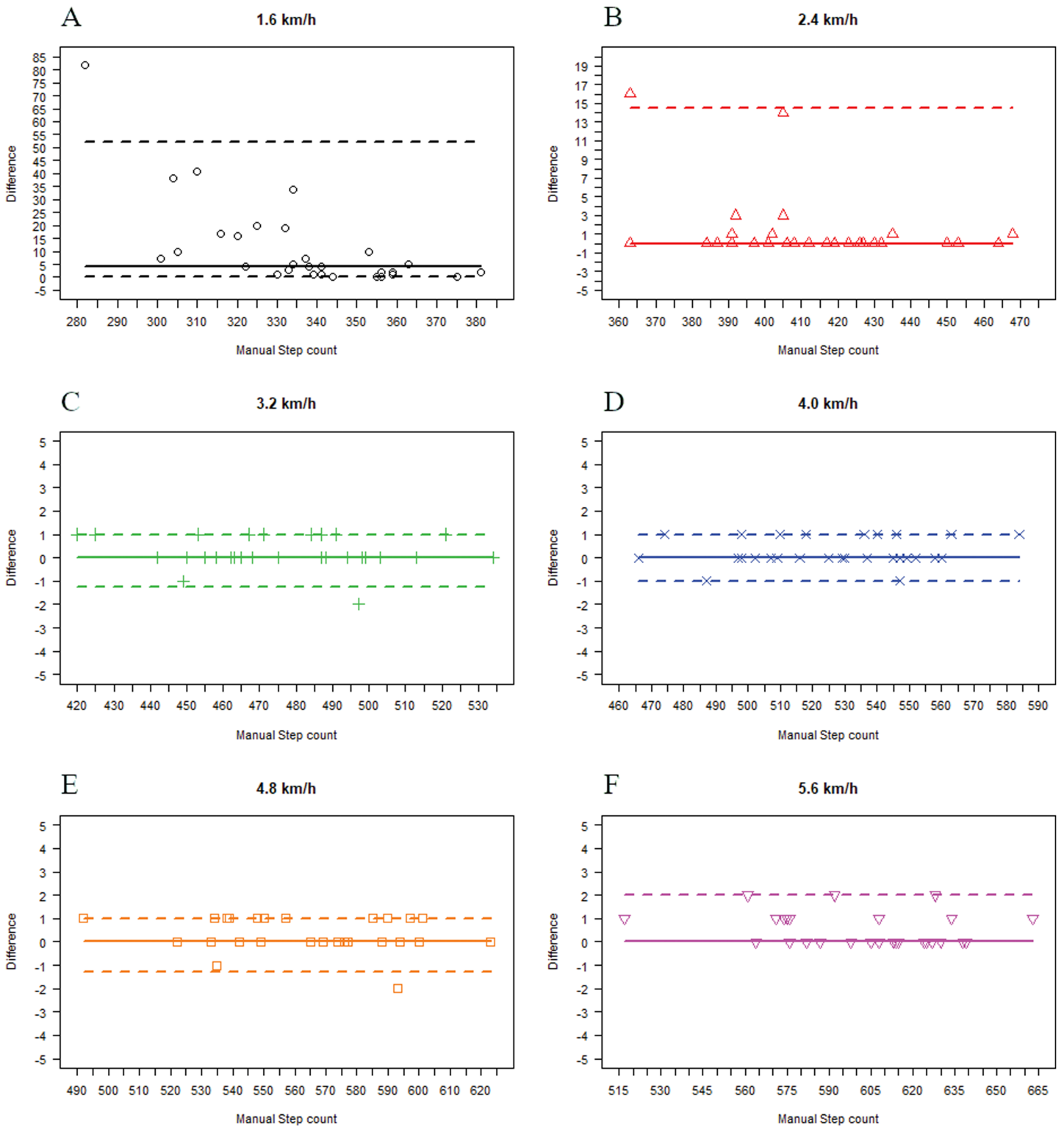
Full-size ⬛ DOI: 10.7717/peerj.9381/fig-3

Figure 4 Bland Altman style plot illustrating the differences between StepWatch™ 3 and manual step count at walking speeds from 1.6 km/h to 5.6 km/h. (A) 1.6 km/h, (B) 2.4 km/h, (C) 3.2 km/h, (D) 4.0 km/h, (E) 4.8 km/h, (F) 5.6 km/h. Full-size ☑ DOI: 10.7717/peerj.9381/fig-4

Despite their tendencies to undercount and overcount steps, the activity trackers may still be clinically relevant at some walking speeds when a cut point of a MAPE ≤3% is used (*Holbrook, Barreira & Kang, 2009*; *Colley et al., 2013*; *Johnson et al., 2015*; *Liu et al., 2015*; *Riel et al., 2016*). The results showed that both Garmin Vivosmart® HR and StepWatch™ 3 miscounted steps at slow walking speeds, and that Garmin Vivosmart® HR miscounted steps at fast walking speeds as well. Taking this into account, we find that Garmin Vivosmart® HR is valid for counting steps at walking speeds from 3.2 km/h to 4.8 km/h and close to the cut point at walking speed of 2.4 km/h with a MAPE of 3.49%. If the aim is to walk 10,000 steps per day, Garmin Vivosmart® HR would miscount 349 daily steps at a walking speed of 2.4 km/h. However, the MAPE also shows that Garmin Vivosmart® HR is far from valid when counting steps at very slow and fast walking speeds. Old people tend to walk slower than younger healthy people and many also reduce their physical activity level (*Burton et al., 2018*). Accurate measuring of walking speed and amount of walking is important because both speed and amount of walking are associated with health outcome in older people (*Graham et al., 2010*; *Lee et al., 2019*). However, at a slow walking speed of 1.6 km/h, Garmin Vivosmart® HR would miscount around ¼ of all steps, which makes it highly unreliable in a clinical setting with patient groups walking at slow speeds.

When assessing MAPE for StepWatch™ 3, we found that this device is more valid in detecting steps than Garmin Vivosmart® HR at all walking speeds. StepWatch™ 3 only exceeds the cut point slightly at a walking speed of 1.6 km/h with a MAPE of 3.60%. This supports previous findings stating that StepWatch™ is one of the most accurate activity trackers for step counting at slow walking speeds (*Lim et al., 2018*). Both devices showed difficulties in counting steps, especially at slow walking speeds, which supports the results of previous studies of activity trackers (*Crouter et al., 2003*; *Steeves et al., 2011*; *Van Remoortel et al., 2012*; *Fortune et al., 2014*; *Fokkema et al., 2017*; *Bunn et al., 2018*; *Gaz et al., 2018*). A plausible explanation suggested in previous studies of other activity trackers may be that the algorithm used for step detection at slow walking speeds was insufficient (*Taraldsen et al., 2011*; *Fortune et al., 2014*) and that very slow walking therefore may not generate sufficient acceleration for registration by accelerometry (*Fokkema et al., 2017*; *Gaz et al., 2018*). It is a challenge that the step-detection algorithm is a closed source in most consumer activity trackers because this prevents analysis of how the accelerometer raw data are processed. Another general issue is that some activity trackers are updated or become outdated before validation studies are published (*Bunn et al., 2018*). It should be noted that Garmin Vivosmart® HR has been discontinued (*Garmin, 2020*). Garmin's newest Vivosmart® is the Garmin Vivosmart® 4, and due to the closed source firmware, it is unknown but also unlikely that newer devices are programed with the same firmware as older models. In general, it should be mentioned that manufacturers frequently update their software. Ongoing optimization and perfection of algorithms are likely to affect activity detection and thereby step counts. Therefore, the results of this study should be interpreted with consideration to the software versions at the time of data collection; 18–27th of April 2016.

The different placements of the activity trackers may also affect their ability to detect steps. Anderson and colleagues found that the placement of the accelerometer affects the criterion validity; that is, ankle-mounted devices seem valid, whereas thigh-worn devices are not considered valid using ActiGraph GT3X for step counting among ward-based adults (*Anderson, Yoward & Green, 2019*). In addition, studies of hip-based activity trackers have found that these devices provide more accurate step counts than wrist-based trackers (*Gaz et al., 2018*). Due to its placement on the wrist, the accelerometer in Garmin Vivosmart® HR will primarily be affected by arm swing and upper body movement, whereas the accelerometer in StepWatch™ 3 is more likely to be affected by shuffling, stride length, etc., due to its placement on the ankle.

As mentioned previously, manual step count is considered the gold standard in laboratory conditions, but measuring daily steps in free-living conditions by manual step count would be very impractical. Results indicate that the research-graded StepWatch™ 3 is the best device for measuring step count in free-living conditions due to its accuracy at both slow and fast walking speeds. However, the cheaper, consumer-graded Garmin Vivosmart® HR can be used in a target population with a narrower range of walking speeds, from relatively slow to moderately fast walking speeds (3.2–4.8 km/h). If the target population meets the criteria for walking speeds of Garmin Vivosmart® HR, using this device could be a cost-benefit-balanced approach to measuring steps. This information could aid decision makers and healthcare personnel when considering to implement physical activity trackers in clinical practice and provide an indication on whether they might consider using a consumer-graded activity tracker instead of a research-graded activity tracker in their target population in relation to walking speed.

Strengths and limitations

The use of different walking speeds under controlled conditions in general and the inclusion of walking speed down to 1.6 km/h in particular are considered a strength of the present study since this has not previously been addressed sufficiently in criterion validity studies. Very slow walking speeds are typically found in older or disabled people; however, the results of the present study cannot be directly compared to older people as only healthy younger adults participated in our study. The gait pattern in younger people may differ from the gait pattern in older people (*Burton et al., 2018*). The present study only investigated step detection under controlled conditions in straight treadmill walking, and it should be noted that sensitivity and specificity may be very different in free-living conditions and in people with disabilities and alternate walking patterns, for instance.

This study adds to the knowledge base on criterion validity of step count under controlled conditions for consumer and research-graded activity trackers. Even though this study investigates the criterion validity of two specific activity trackers, the findings may guide decision makers and healthcare personnel's choice when implementing and applying activity trackers in clinical practice.

Future work

Future studies should validate and compare consumer and research-graded activity trackers in free-living conditions and include non-healthy participants whose movement patterns and mobility disorders might influence the ability of the activity trackers to detect steps (*Steeves et al., 2011*; *Stansfield, Hajarnis & Sudarshan, 2015*; *Fokkema et al., 2017*; *Burton et al., 2018*). Furthermore, it would be relevant to investigate the ability of the activity trackers to detect steps and reject non-steps in activities such as getting in/out of bed, sitting and raising from a chair, jumping, etc., to expand knowledge on the ability of the activity trackers to discriminate steps from other free-living activities in different populations.

CONCLUSION

This study concludes that both the consumer-graded off-the-shelf activity tracker (Garmin Vivosmart® HR) and the research-graded activity tracker (StepWatch™ 3) are valid in detecting steps at selected walking speeds in healthy adults under controlled conditions. However, both activity trackers miscount steps at slow walking speeds, and the consumer-graded off-the-shelf activity tracker also miscounts steps at fast walking speeds. Even though the results reflect the criterion validity of two specific activity trackers under controlled conditions, the findings may inform and support decision makers and healthcare personnel when implementing physical activity trackers and deciding whether to use a consumer-graded or a research-graded activity tracker.

ADDITIONAL INFORMATION AND DECLARATIONS

Author Contributions

- Frederik Rose Svarre conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mads Møller Jensen conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Josephine Nielsen conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Morten Villumsen analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The North Denmark Region Committee on Health Research Ethics stated that ethical approval was not required for this study (Confirmation date 02.16.2016).

Data Availability

The following information was supplied regarding data availability:

Raw data and code are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.9381#supplemental-information.

REFERENCES

Aminian S, Hinckson EA. 2012. Examining the validity of the ActivPAL monitor in measuring posture and ambulatory movement in children. *International Journal of Behavioral Nutrition and Physical Activity* **9(1)**:119 DOI 10.1186/1479-5868-9-119.

Anderson JL, Yoward LS, Green AJ. 2019. A study investigating the validity of an accelerometer in quantification of step count in adult hospital inpatients recovering from critical illness. *Clinical Rehabilitation* **33(5)**:936–942 DOI 10.1177/0269215519829893.

Bergman RJ, Spellman JW, Hall ME, Bergman SM. 2012. Is there a valid app for that? Validity of a free pedometer iPhone application. *Journal of Physical Activity & Health* **9(5)**:670–676 DOI 10.1123/jpah.9.5.670.

Bunn JA, Navalta JW, Fountaine CJ, Reece JD. 2018. Current state of commercial wearable technology in physical activity monitoring 2015–2017. *International Journal of Exercise Science* **11**:503–515.

Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, Howie E. 2018. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. *BMC Geriatrics* **18(1)**:103 DOI 10.1186/s12877-018-0793-4.

Chigateri NG, Kerse N, Wheeler L, MacDonald B, Klenk J. 2018. Validation of an accelerometer for measurement of activity in frail older people. *Gait & Posture* **66**:114–117 DOI 10.1016/j.gaitpost.2018.08.024.

Colley RC, Barnes JD, Leblanc AG, Borghese M, Boyer C, Tremblay MS. 2013. Validity of the SC-StepMX pedometer during treadmill walking and running. *Applied Physiology, Nutrition, and Metabolism* **38(5)**:520–524 DOI 10.1139/apnm-2012-0321.

Crouter SE, Schneider PL, Karabulut M, Bassett DR. 2003. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise* **35(8)**:1455–1460 DOI 10.1249/01.MSS.0000078932.61440.A2.

De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. 2012. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *Journal of Science and Medicine in Sport/Sports Medicine Australia* **15(5)**:418–424 DOI 10.1016/j.jsams.2012.02.002.

Edbrooke L, Lythgo N, Goldsworthy U, Denehy L. 2012. Can an accelerometer-based monitor be used to accurately assess physical activity in a population of survivors of critical illness? *Global Journal of Health Science* **4(3)**:98–107 DOI 10.5539/gjhs.v4n3p98.

Evenson KR, Goto MM, Furberg RD. 2015. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity* **12(1)**:159 DOI 10.1186/s12966-015-0314-1.

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

Fokkema T, Kooiman TJM, Krijnen WP, Van Der Schans CP, De Groot M. 2017. Reliability and validity of ten consumer activity trackers depend on walking speed. *Medicine and Science in Sports and Exercise* **49(4)**:793–800 DOI 10.1249/MSS.0000000000001146.

Fortune E, Lugade VA, Amin S, Kaufman KR. 2015. Step detection using multi-versus single tri-axial accelerometer-based systems. *Physiological Measurement* **36(12)**:2519–2535 DOI 10.1088/0967-3334/36/12/2519.

Fortune E, Lugade V, Morrow M, Kaufman K. 2014. Validity of using tri-axial accelerometers to measure human movement—Part II: step counts at a wide range of gait velocities. *Medical Engineering & Physics* **36(6)**:659–669 DOI 10.1016/j.medengphy.2014.02.006.

Garmin. 2016a. *vivosmart HR | Garmin | Fitness Tracker.* Available at https://buy.garmin.com/en-US/US/into-sports/health-fitness/vivosmart-hr/prod531166.html (accessed 16 May 2016).

Garmin. 2016b. *Sports Features | Garmin Connect.* Available at https://connect.garmin.com/en-US/features/ (accessed 7 May 2016).

Garmin. 2016c. *Garmin owner's: Vivosmart® HR manual.* Olathe: Garmin Ltd.

Garmin. 2020. *Previous Models Fitness Trackers | Garmin.* Available at https://buy.garmin.com/en-US/US/cIntoSports-c571-c10668-p1.html (accessed 12 April 2020).

Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, Henderzahs-Mason JM, Hagen PT. 2018. Determining the validity and accuracy of multiple activity-tracking devices in controlled and free-walking conditions. *American Journal of Health Promotion* **32(8)**:1671–1678 DOI 10.1177/0890117118763273.

Gialamas A, Laurence CO, Yelland LN, Tideman P, Worley P, Shephard MD, Tirimacco R, Willson KJ, Ryan P, Gill J, Thomas DW, Beilby JJ, Members of the PoCT Trial Management Committee. 2010. Assessing agreement between point of care and laboratory results for lipid testing from a clinical perspective. *Clinical Biochemistry* **43(4–5)**:515–518 DOI 10.1016/j.clinbiochem.2009.11.014.

Giannakidou DM, Kambas A, Ageloussis N, Fatouros I, Christoforidis C, Venetsanou F, Douroudos I, Taxildaris K. 2012. The validity of two Omron pedometers during treadmill walking is speed dependent. *European Journal of Applied Physiology* **112(1)**:49–57 DOI 10.1007/s00421-011-1951-y.

GoPro. 2016. GoPro Desktop App - Easily offload and enjoy your GoPro photos and videos, make quick edits and share your favorite shots. *Available at* http://shop.gopro.com/softwareandapp/gopro-app-%7C-desktop/GoPro-Desktop-App.html# (accessed 7 May 2016).

Graham JE, Fisher SR, Bergés I-M, Kuo Y-F, Ostir GV. 2010. Walking speed threshold for classifying walking independence in hospitalized older adults. *Physical Therapy* **90(11)**:1591–1597 DOI 10.2522/ptj.20100018.

Haskell WL, Lee I-M, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD, Bauman A. 2007. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports and Exercise* **39(8)**:1423–1434 DOI 10.1249/mss.0b013e3180616b27.

Hendrick P, Boyd T, Low O, Takarangi K, Paterson M, Claydon L, Milosavljevic S. 2010. Construct validity of RT3 accelerometer: a comparison of level-ground and treadmill walking at self-selected speeds. *Journal of Rehabilitation Research and Development* **47(2)**:157–168 DOI 10.1682/JRRD.2009.04.0047.

Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. 2015. Validity of activity monitor step detection is related to movement patterns. *Journal of Physical Activity & Health* **13(2)**:145–153 DOI 10.1123/jpah.2015-0203.

Holbrook EA, Barreira TV, Kang M. 2009. Validity and reliability of Omron pedometers for prescribed and self-paced walking. *Medicine and Science in Sports and Exercise* **41(3)**:670–674 DOI 10.1249/MSS.0b013e3181886095.

Huang Y, Xu J, Yu B, Shull PB. 2016. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait & Posture* **48**:36–41 DOI 10.1016/j.gaitpost.2016.04.025.

Johnson M, Meltz K, Hart K, Schmudlach M, Clarkson L, Borman K. 2015. Validity of the actical activity monitor for assessing steps and energy expenditure during walking. *Journal of Sports Sciences* **33(8)**:769–776 DOI 10.1080/02640414.2014.964747.

Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, Van der Schans CP, De Groot M. 2015. Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation* **7(1)**:24 DOI 10.1186/s13102-015-0018-5.

Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology* **64(1)**:96–106 DOI 10.1016/j.jclinepi.2010.03.002.

Kumahara H, Ayabe M, Ichibakase M, Tashima A, Chiwata M, Takashi T. 2015. Validity of activity monitors worn at multiple nontraditional locations under controlled and free-living conditions in young adult women. *Applied Physiology, Nutrition, and Metabolism* **40(5)**:448–456 DOI 10.1139/apnm-2014-0183.

Le Masurier GC, Tudor-Locke C. 2003. Comparison of pedometer and accelerometer accuracy under controlled conditions. *Medicine and Science in Sports and Exercise* **35(5)**:867–871 DOI 10.1249/01.MSS.0000064996.63632.10.

Lee I-M, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. 2019. Association of step volume and intensity with all-cause mortality in older women. *JAMA Internal Medicine* **179(8)**:1105 DOI 10.1001/jamainternmed.2019.0899.

Lee JA, Williams SM, Brown DD, Laurson KR. 2015. 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. *Journal of Sports Sciences* **33(10)**:991–1000 DOI 10.1080/02640414.2014.981848.

Lim SER, Ibrahim K, Sayer AA, Roberts HC. 2018. Assessment of physical activity of hospitalised older adults: a systematic review. *Journal of Nutrition, Health & Aging* **22(3)**:377–386 DOI 10.1007/s12603-017-0931-2.

Liu S, Brooks D, Thomas S, Eysenbach G, Nolan RP. 2015. Lifesource XL-18 pedometer for measuring steps under controlled and free-living conditions. *Journal of Sports Sciences* **33(10)**:1001–1006 DOI 10.1080/02640414.2014.981847.

Modus Health. 2014. *StepWatch™ 3 activity monitor & USB docking station—Version 3.4 user manual*. Washington, D.C.: Modus Health.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. 2010a. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* **19(4)**:539–549 DOI 10.1007/s11136-010-9606-8.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. 2010b. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology* **63(7)**:737–745 DOI 10.1016/j.jclinepi.2010.02.006.

Riel H, Rathleff CR, Kalstrup PM, Madsen NK, Pedersen ES, Pape-Haugaard LB, Villumsen M. 2016. Comparison between mother, ActiGraph wGT3X-BT, and a hand tally for measuring steps at various walking speeds under controlled conditions. *PeerJ* **4(1)**:e2799 DOI 10.7717/peerj.2799.

Rosenkranz RR, Rosenkranz SK, Weber C. 2011. Validity of the Actical accelerometer step-count function in children. *Pediatric Exercise Science* **23(3)**:355–365 DOI 10.1123/pes.23.3.355.

Ryan CG, Grant PM, Tigbe WW, Granat MH. 2006. The validity and reliability of a novel activity monitor as a measure of walking. *British Journal of Sports Medicine* **40(9)**:779–784 DOI 10.1136/bjsm.2006.027276.

Sellers C, Dall P, Grant M, Stansfield B. 2016. Validity and reliability of the activPAL3 for measuring posture and stepping in adults and young people. *Gait & Posture* **43**:42–47 DOI 10.1016/j.gaitpost.2015.10.020.

Shephard RJ. 1988. PAR-Q, Canadian home fitness test and exercise screening alternatives. *Sports Medicine* **5(3)**:185–195 DOI 10.2165/00007256-198805030-00005.

Stansfield B, Hajarnis M, Sudarshan R. 2015. Characteristics of very slow stepping in healthy adults and validity of the activPAL3™ activity monitor in detecting these steps. *Medical Engineering & Physics* **37(1)**:42–47 DOI 10.1016/j.medengphy.2014.10.003.

Steeves JA, Tyo BM, Connolly CP, Gregory DA, Stark NA, Bassett DR. 2011. Validity and reliability of the Omron HJ-303 tri-axial accelerometer-based pedometer. *Journal of Physical Activity & Health* **8(7)**:1014–1020 DOI 10.1123/jpah.8.7.1014.

Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. 2014. Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport/Sports Medicine Australia* **17(5)**:496–500 DOI 10.1016/j.jsams.2013.10.241.

Taraldsen K, Askim T, Sletvold O, Einarsen E, Bjåstad K, Indredavik B, Helbostad J. 2011. Evaluation of a body-worn sensor system to measure physical activity in older people with impaired function. *Physical Therapy* **91(2)**:277–285 DOI 10.2522/ptj.20100159.

Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. 2019. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. *PLOS ONE* **14(5)**:e0216891 DOI 10.1371/journal.pone.0216891.

Thompson WR. 2015. Worldwide survey of fitness trends for 2016: 10th Anniversary Edition. *ACSM's Health & Fitness Journal* **19(6)**:9–18 DOI 10.1249/FIT.0000000000000164.

Thompson WR. 2016. Worldwide survey of fitness trends for 2017. *ACSM's Health & Fitness Journal* **20(6)**:8–17 DOI 10.1249/FIT.0000000000000252.

Thompson WR. 2018. Worldwide survey of fitness trends for 2019. *ACSM's Health & Fitness Journal* **22(6)**:10–17 DOI 10.1249/FIT.0000000000000438.

Thompson WR. 2019. Worldwide survey of fitness trends for 2020. *ACSM's Health & Fitness Journal* **23(6)**:10–18 DOI 10.1249/FIT.0000000000000526.

Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. 2017. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. *Physical Therapy* **97(5)**:581–588 DOI 10.1093/ptj/pzx010.

Tudor-Locke C, Bassett DR, Shipe MF, McClain JJ. 2011a. Pedometry methods for assessing free-living adults. *Journal of Physical Activity & Health* **8(3)**:445–453 DOI 10.1123/jpah.8.3.445.

Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, Hatano Y, Inoue S, Matsudo SM, Mutrie N, Oppert J-M, Rowe DA, Schmidt MD, Schofield GM, Spence JC, Teixeira PJ, Tully MA, Blair SN. 2011b. How many steps/day are enough? For adults. *International Journal of Behavioral Nutrition and Physical Activity* **8(1)**:79 DOI 10.1186/1479-5868-8-79.

Tudor-Locke C, Williams JE, Reis JP, Pluto D. 2002. Utility of pedometers for assessing physical activity: convergent validity. *Sports Medicine* **32(12)**:795–808 DOI 10.2165/00007256-200232120-00004.

Van Remoortel H, Giavedoni S, Raste Y, Burtin C, Louvaris Z, Gimeno-Santos E, Langer D, Glendenning A, Hopkinson NS, Vogiatzis I, Peterson BT, Wilson F, Mann B, Rabinovich R, Puhan MA, Troosters T, PROactive consortium. 2012. Validity of activity monitors in health and chronic disease: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity* **9(1)**:84 DOI 10.1186/1479-5868-9-84.

zebris Medical GmbH. 2015. *FDM-T technical specifications and user manual.* Isny im Allgä: zebris Medical GmbH.