

# Original Research

## Validity and Reliability Evaluation of Four Commercial Activity Trackers' Step Counting Performance

Katrine Hauer Tophøj, MSc,<sup>1</sup> Matilde Grøndahl Petersen, MSc,<sup>1</sup> Casper Sæbye, MD,<sup>2</sup> Thomas Baad-Hansen, PhD, MD,<sup>2</sup> and Stefan Wagner, PhD<sup>1</sup>

<sup>1</sup>Department of Engineering, Aarhus University, Aarhus, Denmark.

<sup>2</sup>Department of Orthopedic Surgery, Aarhus University Hospital, Aarhus, Denmark.

### Abstract

**Introduction:** Activity trackers are useful tools for physical rehabilitation purposes. Most available activity trackers are designed for fitness and wellness use, lacking in both accuracy and precision at lower speeds. Validity and reliability at all clinically relevant speeds are crucial selection criteria for use in clinical practice. The aim of this study was to assess the validity and reliability of four consumer-based activity trackers at clinical relevant walking speeds for patient groups undergoing rehabilitation.

**Methods:** The four commercial activity trackers Fitbit Surge (FS), Fitbit Charge HR (FC), Microsoft Band 2 (MB), and A&D 101NFC Activity Monitor (A&D) were evaluated at 2, 4, 4.5, and 5.5 km/h. Twenty healthy participants aged  $25.6 \pm 2$  years walked on a treadmill at the four velocities in two trials of 100 steps each. Mean average percentage error (MAPE), intraclass correlation coefficient (ICC), and Bland-Altman 95% limits of agreement were calculated to assess validity and reliability.

**Results:** MAPE levels were between  $-8\%$  and  $-6\%$  for FS,  $-15\%$  and  $0\%$  for MB,  $7\%$  and  $21\%$  for FC, and  $-53\%$  and  $1\%$  for AD. The biggest inaccuracies were seen at 2 km/h, where AD underestimated by 53%. The highest accuracy was predominantly found with MB and AD, which overestimated  $\leq 2\%$  at velocities  $\geq 4$  km/h. ICC was moderate (0.73) for FS, good (0.88) for MB, moderate (0.52) for FC, and excellent (0.98) for AD.

**Conclusions:** MB, FS, and AD accurately counted steps, when participants walked with velocities corresponding to a brisk walk ( $\geq 4$  km/h). Walking at lower speeds ( $\leq 2$  km/h) was not counted accurately. Thus, the four evaluated activity trackers are not useful for patient groups walking at lower speeds during rehabilitation, nor for counting indoor walking.

**Keywords:** activity tracker, step counting, pedometer, rehabilitation, telehealth, telemedicine, telerehabilitation, e-health, home health monitoring, wearable sensors, physical activity, soft tissue sarcoma

### Introduction

Step counting using commercial activity trackers has become the most used metric to measure physical activity, because it is an easy and low-cost method.<sup>1,2</sup> In general, activity trackers accommodate the need for quantified feedback, which has shown to increase the physical activity.<sup>1</sup> The interest in physical activity has increased in clinical practice, where it has become an important part of treatment and rehabilitation regarding numerous diseases.<sup>1</sup> Commercial activity trackers are widely recognized for their estimation of physical activity; however, they are often not scientifically validated. Valid and reliable performance of activity trackers is crucial if activity trackers are to be successfully adopted by clinicians.<sup>1</sup> Valid step counting would allow the clinicians to gain insight into the physical activity levels for patients performing self-monitored home rehabilitation. This sharing of information would be beneficial for both clinicians and patients in situations where ambulation is essential for rehabilitation.<sup>3</sup>

Patients, for whom unsupervised home exercises are required as part of the rehabilitation intervention, may not adhere to the rehabilitation regimen prescribed by the clinicians, and with current state-of-the-art, most clinicians have no means of verifying patient adherence levels. An implementation of a wearable activity tracker as part of the rehabilitation process, for example, as part of a tele-rehabilitation system, would make it possible for both the clinician and the patient to assess if the patient carries out the recommended training based on an objective measure. The clinicians can simultaneously assess the patient's ability to act on deviations from the recommended training and intervene if there are problems.

The ability to measure steps at slow walking speed has been a focus area in clinical research, because patients often have reduced walking velocity. Research-grade activity trackers have

shown significantly more valid results at slow walking speeds ( $\leq 2$  km/h)<sup>4</sup> compared with especially older consumer-based trackers with mechanical accelerometers, which have been reported unable to deliver valid and reliable data.<sup>5</sup> However, procurement costs for research-grade trackers compared with consumer-grade trackers are high, and the research-grade trackers are not always designed to be user friendly<sup>2</sup> and thus providing practicality problems for clinical usage scenarios. On the contrary, the competitive market of consumer-grade activity trackers delivers affordable and intuitive trackers.

To implement activity trackers in clinical research, including for telerehabilitation interventions, the first step is to find a valid and reliable activity tracker, which can safely monitor physical activity (steps) at the specific walking velocities of the target group of the intervention. Different target groups have different walking velocities during rehabilitation, but most share the common trait of having a slower walking speed than healthy individuals.

The clinically relevant velocities selected for this study were based on the maximum walking speeds of postoperative soft tissue sarcoma (STS) patients who have undergone surgery in the lower extremities. STS is a rare type of cancer where surgery is the mainstay of treatment.<sup>6</sup> However, the results are of general interest to all patient groups with reduced walking speeds.

The aim of this study was to investigate the validity and reliability of four activity trackers (Fitbit Surge [FS], Fitbit Charge HR [FC], Microsoft Band 2 [MB], and A&D 101NFC Activity Monitor [A&D]) at clinically relevant velocities for patient groups undergoing rehabilitation.

## Materials and Methods

To evaluate the four activity trackers, this study was conducted in a research laboratory at Aarhus University,

where the participants walked on a treadmill at four different speeds. The participants walked 100 steps twice at each velocity for comparison.

## PARTICIPANTS

Twenty healthy voluntary participants were recruited from the student population at Aarhus University.

## DEVICES

The evaluated activity trackers were FS, MB, FC, and A&D. Detailed information about the examined trackers is listed in *Table 1* to get an overview of the specification, function, and price. All devices were based on a 3-axis accelerometer technology.

## WALKING SPEEDS

An evaluation of 6-min walking test of 13 postsurgical STS patients at Aarhus University Hospital<sup>7</sup> showed that 75% of the patients walked with a velocity between 4 and 5.5 km/h. Based on these observations, 4, 4.5, and 5.5 km/h were chosen as clinically relevant walking speeds to represent a brisk walk.

Walking speeds based on 6-min walking tests do not represent habitual walking speed and small indoor daily activities, which could also be of interest when evaluating patient progress in a rehabilitation intervention. Previous studies have estimated walking speed related to indoor daily activities to lie in the interval of 2–3 km/h.<sup>8,9</sup> In this study, 2 km/h was chosen to represent the walking speed for slow indoor walking.

## PROTOCOL

Each patient's weight and height were measured and entered into the user profile of the MB and Fitbit devices to

**Table 1. Information About the Examined Activity Trackers**

DEVICE	TYPE	PLACEMENT	SENSORS	SOFTWARE	BATTERY LIFE	PRICE (\$)
FS	Accelerometer	Nondominant wrist	GPS, 3-axis accelerometer, gyroscope, compass, optical heart rate, altimeter, ambient light sensor	Fitbit for iOS and android	7 days (GPS on 10 h)	200
FC	Accelerometer	Nondominant wrist	3-axis accelerometer, optical heart rate, altimeter	Fitbit for iOS and android	7 days	100
MB	Accelerometer	Nondominant wrist	Optical heart rate sensor, 3-axis accelerometer, gyroscope, GPS, ambient light sensor, skin temperature sensor, UV sensor, capacitive sensor, galvanic skin response, microphone, barometer	Microsoft Health v 1.3.21021.1	48 h	220
A&D	Accelerometer	Hip	3-axis accelerometer	—	>1 year	80

A&D, A&D 101NFC activity monitor; FC, Fitbit Charge HR; FS, Fitbit Surge; MB, Microsoft Band 2.

calibrate the two trackers. The participant wore four activity trackers simultaneously: three watch-style trackers on their nondominant wrist and the A&D tracker on the hip. The participants wore their daily clothes and shoes. The watch-style trackers needed to move continuously for about 10–15 steps to start counting and updating in real time according to the user’s guide,<sup>10</sup> which is why the participants began with a 1–2-min warm-up on the treadmill before the test began. After the warm-up, the test participants walked intervals of 100 steps at the four different walking speeds: 2, 4, 4.5, and 5.5 km/h. Each participant carried out the test two times at each speed and walked 800 steps in total. The steps were manually counted by the facilitator and the step counts on the trackers were read in-between every interval.

STATISTICAL ANALYSIS

Descriptive analyses were conducted for the participants and for each device. The steps measured by the trackers were tested for normality with the Shapiro–Wilk normality test. If the data were normally distributed, validity was assessed with the mean average percentage error (MAPE) for each device and velocity, and if not, the median average percentage error was used. The MAPE was calculated with the following equation for each device and velocity:

$$MAPE = \frac{\text{average mean} - 100 \text{ steps}}{100 \text{ steps}} \times 100\%$$

The intrarate reliability was assessed by intraclass correlation coefficient (ICC) (two-way random-effects models: ICC [2,1]) between the first and second trial at each velocity to investigate the reliability of each device, and thus, the reproducibility for the tracker to measure the same step count for the same person and velocity each time. Less than 0.5 was interpreted as poor, 0.5–0.74 as moderate, 0.75–0.89 as good, and 0.9–1.0 as excellent.<sup>11</sup> Negative ICC values can occur in statistical programs, but are not theoretically possible. Therefore, negative ICC values will not be interpreted.<sup>12</sup>

A Bland–Altman plot with limits of agreement was used to determine the error of measurement of each tracker. STATA/IC 14.2 was used as the statistical program.

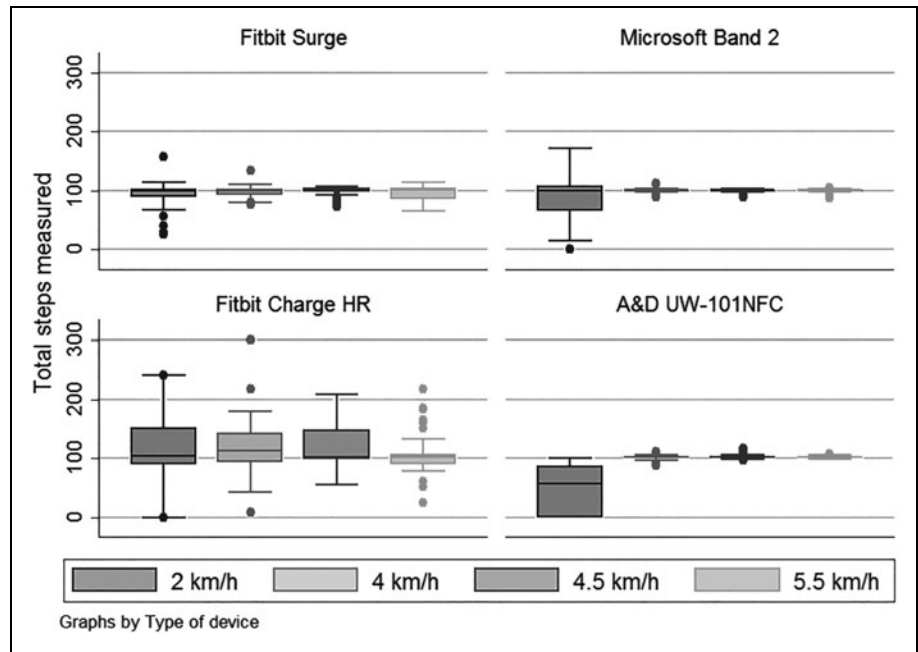


Fig. 1. Box plot of the step measurements recorded by the FS, the MB, the FC, and the A&D at treadmill walking speeds of 2, 4, 4.5, and 5.5 km/h. The length of the box depicts the interquartile range. A&D, A&D UW-101NFC activity monitor; FC, Fitbit Charge HR; FS, Fitbit Surge; MB, Microsoft Band 2.

Results

The participating volunteers in the study included 20 persons (10 males and 10 females) aged 25.6 ± 2 years.

A graphical representation of the results is depicted in the box plot of Figure 1.

The results of the mean percentage error calculations are presented in Table 2. The results are presented individually for the different activity trackers (FS, MB, FC, and A&D) at 2, 4, 4.5, and 5.5 km/h (Speed), respectively, as the MAPE, standard deviation (SD), 95% confidence interval, and the lowest and highest percentage measurement error (Min and Max). A negative coefficient indicates underestimation and a positive coefficient indicates overestimation.

VALIDITY

At 2 km/h, it was observed that A&D made the biggest underestimation of steps compared with the other activity trackers with a mean average error rate of, respectively, –53% for A&D, –8% for FS, –15% for MB, and 15% for FC. FS was the most accurate at this speed considering both MAPE and the SD, which was 23 steps compared with 46, 56, and 42 steps for MB, FC, and A&D, respectively.

At 4 km/h, the smallest percentage errors were observed with MB (0%) and A&D (1%). The error rate for FS and FC were –2% and 18%, respectively. The largest percentage error was seen by FC. The SD for all activity trackers was smaller

**Table 2. Mean Average Percentage Error, Standard Deviation, Confidence Interval 95%, and the Lowest and Highest Percentage Measurement Error (Min and Max) for Each Device and Speed**

	SPEED (KM/H)	MAPE (%)	SD (%)	95% CI		MIN (%)	MAX (%)
				LOWER	UPPER		
FS	2	-8	23	-15	0	-73	57
	4	-2	9	-5	1	-24	34
	4.5	-2	8	-5	1	-27	8
	5.5	-6	13	-10	-2	-34	15
MB	2	-15	46	-29	1	-100	72
	4	0	4	-1	2	-11	12
	4.5	0	3	-1	1	-10	4
	5.5	0	3	-1	1	-12	6
FC	2	15	56	-3	33	-100	140
	4	18	50	2	34	-91	200
	4.5	21	36	10	33	-45	108
	5.5	7	36	-5	18	-76	118
A&D	2	-53	42	-67	-40	-100	1
	4	1	4	-1	2	-13	11
	4.5	2	3	1	3	-4	17
	5.5	2	2	2	3	-1	7

CI, confidence interval; MAPE, mean average percentage error; SD, standard deviation.

compared with the SDs at 2 km/h. The lowest SDs were observed with MB and A&D with an SD of ±4%.

At 4.5 km/h, the percent errors were -2%, 0%, 21%, and 2% for FS, MB, FC, and A&D, respectively. FC was the only activity tracker that did not measure within ±2%.

At 5.5 km/h, the lowest percent errors were observed with MB (0%) and A&D (2%). FS underestimated 6% and FC overestimated 7%.

**RELIABILITY**

The reliability of the activity trackers is expressed by the ICC for each activity tracker at each rate (Table 3).

A&D had a total ICC score of 0.96, which is interpreted as excellent consistency between the first and second measurement.<sup>11</sup> MB had a good consistency, FS had moderate consistency, and FC showed poor consistency. At 2 km/h the ICC varied from poor (0.51) for FS to excellent (0.93) for A&D.

**Table 3. Intraclass Correlation Coefficient (Corresponding 95% Confidence Interval)**

	2 KM/H	4 KM/H	4.5 KM/H	5.5 KM/H	TOTAL
FS	0.51 (0.11; 0.77)	0.50 (0.08; 0.77)	0.73 (0.44; 0.89)	0.78 <sup>a</sup> (0.52; 0.91)	0.58 (0.41; 0.70)
MB	0.79 <sup>a</sup> (0.55; 0.91)	<i>-0.21 (-0.63; 0.27)</i>	<i>0.06 (-0.41; 0.49)</i>	<i>0.44 (-0.01; 0.73)</i>	0.79 <sup>a</sup> (0.69; 0.86)
FC	0.52 (0.07; 0.79)	<i>0.31 (-0.16; 0.66)</i>	0.43 (0.01; 0.73)	<i>0.06 (-0.40; 0.48)</i>	0.35 (0.14; 0.53)
A&D	0.93 <sup>b</sup> (0.83; 0.97)	0.60 (0.23; 0.82)	<i>0.36 (-0.08; 0.68)</i>	0.82 <sup>a</sup> (0.59; 0.92)	0.96 <sup>b</sup> (0.95; 0.98)

<sup>a</sup>Good.

<sup>b</sup>Excellent.

*Italic represent negative values, not interpretable intraclass correlation coefficient values.*

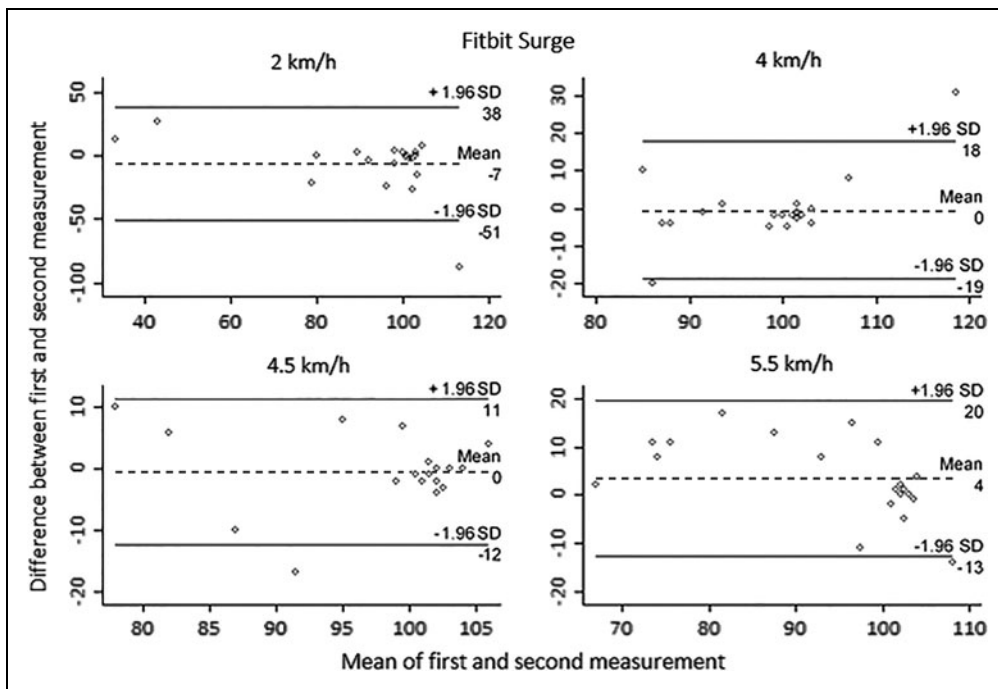


Fig. 2. Bland-Altman plot for FS at each velocity for first and second measurement. Limits of agreement are depicted with red lines. Mean difference is depicted with a dashed line.

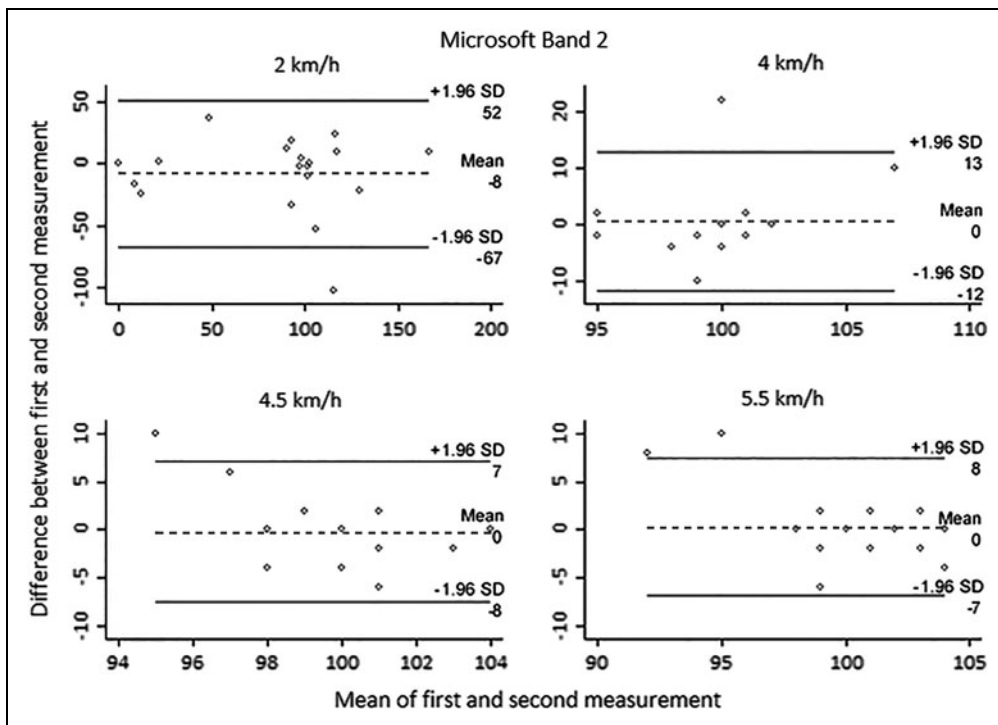


Fig. 3. Bland-Altman plot for MB at each velocity for first and second measurement. Limits of agreement are depicted with red lines. Mean difference is depicted with a dashed line.

The Bland-Altman plot illustrates the agreement between first and second measurements for each speed and activity tracker. Each plot contains 20 data points, one for each participant (Figs. 2-5).

Table 4 shows the values from the Bland-Altman plot with mean difference between first and second measurement, 95% limits of agreement (LOA), the difference between lower and upper LOA, and a *p*-value for a paired *t*-test. At the Bland-Altman plot, all activity trackers tended to have a cluster around 100 steps for mean of the measurements (*x*-axis) and about 0 in difference (*y*-axis) at velocities  $\geq 2$  km/h, which is also the ground truth. However, FC generally had a greater variation from these trends (Fig. 4). The lowest agreement was generally seen at 2 km/h in the Bland-Altman plots, where the mean differences were furthest from 0 (ground truth) with -7, -8, -34, and -2 steps (Table 4) for FS (Fig. 2), MB (Fig. 3), FC (Fig. 4), and A&D, respectively (Fig. 5). MB had a mean of 0 between first and second measurement at velocities  $>2$  km/h and the corresponding limits of the agreement interval became smaller as the speed became faster (Fig. 3 and Table 4). MB had the smallest range of agreement (15 steps) at 4.5 km/h. A&D had the lowest range in limits of agreement at speeds of 2, 4, and 5.5 km/h (65, 16, and 4, respectively) (Table 4). The highest limits of agreement ranges were FC (186, 233, 153, and 193 steps) as depicted in the Bland-Altman plots (Fig. 4). Statistically significant differences were only observed in the measurement points from the test-retest for FC at 2 km/h (Table 4) evaluated by a paired *t*-test.

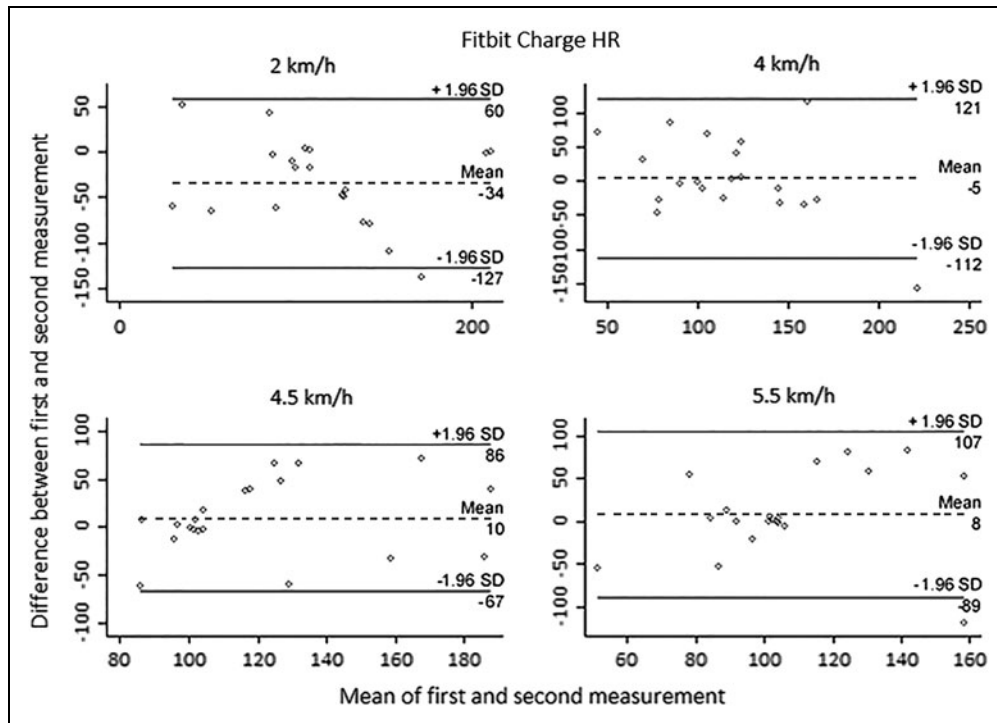


Fig. 4. Bland–Altman plot for FC at each velocity for first and second measurement. Limits of agreement are depicted with red lines. Mean difference is depicted with a dashed line.

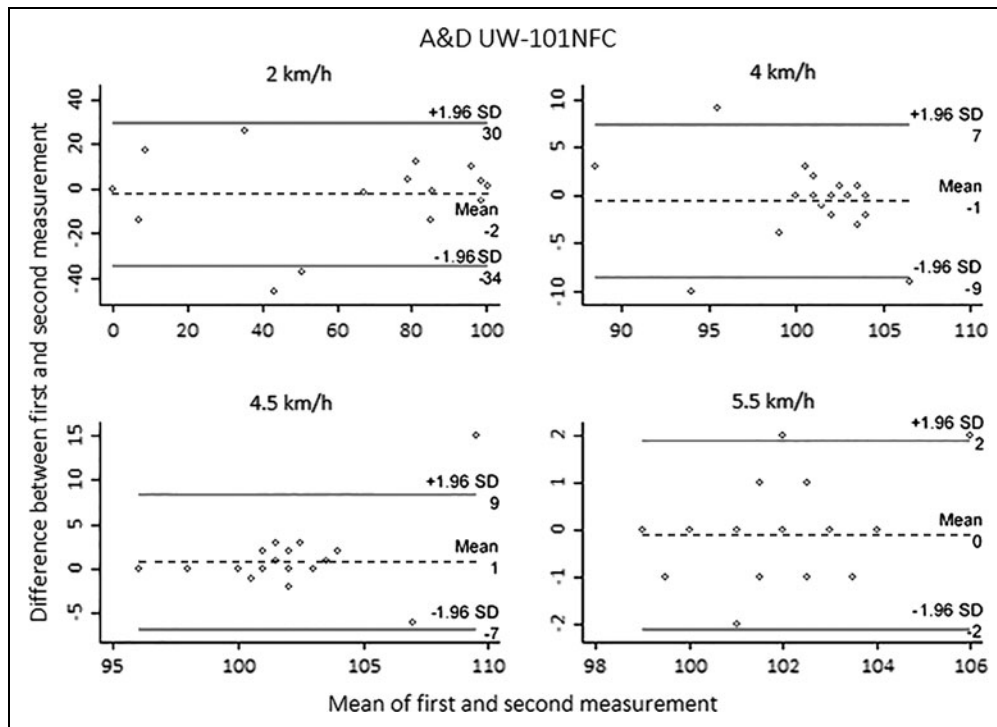


Fig. 5. Bland–Altman plot for A&D UW-101NFC at each velocity for first and second measurement. Limits of agreement are depicted with red lines. Mean difference is depicted with a dashed line.

## Discussion

The validity and test–retest reliability performance were tested for four commercial activity trackers at the walking speed of postoperative sarcoma patients. The validity increased in general for all trackers when the walking speed increased. Considerable variability was found among the different trackers in the MAPE (from –53% to 0%). FS, MB, and A&D have percentage errors  $\leq 2\%$  when the walking velocity was  $\geq 4$  km/h. A slight systematic overestimation of steps (1–2 steps) was seen, which might be due to the test design, when stepping on and off the treadmill. FS generally showed the most accurate results of the tested activity trackers at 2 km/h. MB was very close to FS, but the percentage error was large at 2 km/h and is therefore not considered very accurate at this velocity.

The A&D was the most accurate at velocities  $\geq 4$  km/h and performed badly at 2 km/h. This might be due to the measurement range of the accelerometer used and/or the algorithm used to determine when to count movement as a step. A&D would be suitable for registration of brisk walking ( $>4$  km/h) where it showed  $<3\%$  inaccuracy in step counting.

FC had the biggest deviance at all speeds, and it poses a different challenge than A&D. FC had a mean step count at 115 steps at 2 km/h, but the precision (ICC) was moderate at 0.52. This might be due to the study design, where some discrepancies were observed. Discrepancies included that the FC was very slow to update steps on the display and was therefore challenged in relation to updating before the next 100



**Table 4. Bland–Altman 95% Limits of Agreement**

	SPEED (KM/H)	MEAN DIF.	LOA 95%		LOA 95% DIF	P
			LOWER	UPPER		
FS	2	-7	-51	38	89	0.2027
	4	-0	-19	18	37	0.8325
	4.5	-0	-12	11	24	0.7430
	5.5	4	-13	20	33	0.0696
MB	2	-8	-67	52	118	0.2760
	4	0	-12	13	25	0.7246
	4.5	0	-8	7	15	0.8154
	5.5	0	-7	8	15	0.7199
FC	2	-34	-127	60	186	0.0045 <sup>a</sup>
	4	5	-112	121	233	0.7138
	4.5	10	-67	86	153	0.2775
	5.5	8	-89	107	196	0.4625
A&D	2	-2	-34	30	65	0.5351
	4	-1	-9	7	16	0.5529
	4.5	1	-7	9	15	0.3124
	5.5	0	-2	2	4	0.6663

<sup>a</sup>Statistical significant difference.  
LOA, limits of agreement.

Downloaded by Gothenburg University Library from online.liebertpub.com at 01/26/18. For personal use only.

steps were taken. To accommodate this problem, the facilitator waited for about 30 s and checked several times, if the step count changed before the next session started. The observations regarding update did not match the user’s guide, which said that it began to live update after 10–15 steps. This issue could have led to worse results for the FC than reasonable. To the authors’ knowledge, no previous study has addressed the issue with slow step update. Future research should consider longer measuring periods and to wait longer time before readings are saved. Our results correspond to previous study results where FC had an error margin of 26% when walking at 2 km/h.<sup>13</sup> No previous studies have been found to compare these trackers’ validity or reliability concerning step detection when walking on a treadmill.

When comparing the reliability for all velocities, it was found that A&D had an excellent total ICC score, MB was good, FS was moderate, and FC had a poor ICC. The ICC values were negative at some velocities, which is not

theoretically possible.<sup>12</sup> This can happen when the difference between the measurements for the participants is smaller than the difference within the measurements for the same participant and when the numbers of raters are small. In this study, both scenarios are true, because all participants walked the same number of steps, so the difference between participants was small and the numbers of raters were two, because the same activity tracker measured the same person twice.

From a clinical perspective, validity and reliability are important characteristics of any measurement method, as clinicians make decisions based on these measurements. Previous studies have estimated that an acceptable average measurement error under controlled conditions is within  $\pm 3\%$ .<sup>14</sup> Based on this, MB and A&D are clinically relevant at speeds  $\geq 4$  km/h, and FS is clinically relevant at 4 and 4.5 km/h. However, there are no activity trackers that can be considered clinically relevant at all speeds.

The change in activity levels over time could also be of clinical interest during rehabilitation. Thus, when looking at change over time, it is interesting whether repeated measurements match. Measurement methods used for individual clinical assessment should have an ICC score  $\geq 0.9$ .<sup>15</sup> In this context, only A&D, which has a total ICC of 0.96, is assessed to be clinically relevant.

The inter-reliability of a Fitbit Ultra device has previously been shown to have good correlation between minute to minute and hour to hour, and excellent correlation by day to day.<sup>16</sup>

Our study demonstrated that validity and reliability of activity trackers depended on both the type of activity tracker and the walking speed. The finding that the quality of measurement increased as walking speed increased is in line with previous studies that have examined the quality of other activity trackers.<sup>5,13,17–19</sup>

Healthcare professionals currently make decisions based on the patient’s subjective assessment of physical activity.<sup>20</sup> The patient’s subjective assessment has previously been shown to deviate up to 500% compared to the objective measurements recorded by activity trackers.<sup>20</sup> Thus, one could argue that although the activity trackers are not always completely precise and reliable in all usage scenarios, they provide a more valid overview of the patient’s actual activity level than can be obtained through the patient’s subjective assessments.

Manual counting of steps was chosen as the gold standard method in this study, and it was estimated that this method had an error margin of zero to three steps due to facilitator error. This error margin was assessed based on the possibility

that additional steps could be taken when the subject went on and off the treadmill. It can therefore be argued that the FS cannot be rejected to have clinical relevance at speeds  $\geq 4$  km/h. An average measurement error of  $\pm 3\%$  is a narrow error margin, given that a small measurement error of physical activity cannot cause fatal consequences for the person's health. Other studies recommend that the average measurement error should be  $< 20\%$  when activity meters are used in uncontrolled environments.<sup>14</sup> In this context, both FS and MB are clinically relevant at all speeds, and A&D is clinically relevant at speeds  $\geq 4$  km/h.

The approach with using a treadmill was advantageous in controlling the different walking speeds. However, using a treadmill has previously been reported to possibly affect the natural gait and thus influence the activity trackers. Changes include higher cadence and shorter stride.<sup>5</sup> Some participants felt walking in slow pace to be unnatural and noted that their gait pattern might have changed compared with free-walking conditions. This was in some cases reinforced by being aware of the unfamiliar gait speeds. High accuracy in steps on treadmill has previously been found to correlate poorly with high accuracy in free walking.<sup>21</sup> Therefore, precautions must be made when transferring the results to free-walking conditions. However, using a treadmill design is considered to be an acceptable evaluation method.<sup>5</sup>

The study population was younger than the intended patient target groups, including sarcoma patients undergoing rehabilitation. This is a common study limitation in most related work.<sup>5</sup> Thus, these experiments should be further evaluated with relevant target populations under naturalistic conditions in the future.

Finally, we recommend that existing commercial activity tracker software should be modified for clinical usage to fit into current clinical practice of hospitals and outpatient clinics. In this study, a range of practicality issues exist that are not currently met by the vendors of any of the evaluated activity trackers, especially if the activity trackers are to be used as part of a telerehabilitation intervention. These include anonymized tools for controlling devices being used by multiple users over time, and more easy access to the data for clinical support systems.

## Conclusions

Step counting accuracy of four commercial activity trackers was evaluated in a treadmill-based experiment: FS, FC, MB, and A&D. Poor accuracy was seen at walking speeds of 2 km/h, while the quality increased at velocities  $\geq 4$  km/h. None of the assessed activity trackers could be concluded to

be clinically valid and reliable at all speeds. This poses a challenge in relation to the use of activity trackers in a clinical context, especially in the early rehabilitation phase, where low walking speeds can be expected. Thus, activity trackers are not useful for patient groups walking at lower speeds during rehabilitation or for counting indoor walking at these speeds.

## Author Disclosure Statement

No competing financial interests exist.

## REFERENCES

1. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: A systematic review. *JAMA* 2007;298:2296–2304.
2. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;12:159.
3. Singh AK, Farmer C, Van Den Berg ML, Killington M, Barr CJ. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. *Disabil Health J* 2016;9:320–323.
4. Harrison SL, Horton EJ, Smith R, Sandland CJ, Steiner MC, Morgan MD, et al. Physical activity monitoring: Addressing the difficulties of accurately detecting slow walking speeds. *Heart Lung* 2013;42:361–364.e1.
5. Beevi FH, Miranda J, Pedersen CF, Wagner S. An evaluation of commercial pedometers for monitoring slow walking speed populations. *Telemed J E Health* 2016;22:441–449.
6. ESMO/European Sarcoma Network Working Group. Soft tissue and visceral sarcomas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2014;25 Suppl 3:iii102–iii112.
7. Sæbye, C. PhD thesis: Functional outcome and quality of life after limb-sparing surgery for soft tissue sarcoma. In preparation 2018
8. Hayes TL, Hagler S, Austin D, Kaye J, Pavel M. Unobtrusive assessment of walking speed in the home using inexpensive PIR sensors. *Conf Proc IEEE Eng Med Biol Soc* 2009;2009:7248–7251.
9. Stone E, Skubic M, Rantz M, Abbott C, Miller S. Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders. *Gait Posture* 2015;41:57–62.
10. Fitbit, Inc. How accurate are Fitbit trackers?. December 10 2016; Available at [https://help.fitbit.com/articles/en\\_US/Help\\_article/1136#clipbased](https://help.fitbit.com/articles/en_US/Help_article/1136#clipbased) (last accessed May 20, 2017).
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–163.
12. Giraudeau B. Negative values of the intraclass correlation coefficient are not theoretically possible. *J Clin Epidemiol* 1996;49:1205–1206.
13. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the future patient trial, phase I. *Sensors (Basel)* 2017;17:pii: E211.
14. Thorup CB, Andreasen JJ, Sorensen EE, Gronkjaer M, Dinesen BI, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. *BMJ Open* 2017;7:e011742.
15. Rehabilitation Institute of Chicago, Center for Rehabilitation Outcomes Research, Northwestern University Feinberg School of Medicine,



Department of Medical Social Sciences Informatics group. Rehabilitation Measures Database. 2012. Available at [www.rehabmeasures.org/rehabweb/rhstats.aspx](http://www.rehabmeasures.org/rehabweb/rhstats.aspx) (last accessed June, 4, 2017).

16. Dontje ML, de Groot M, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the Fitbit activity tracker: An inter-device reliability study. *J Med Eng Technol* **2015**;39:286–290.
17. Fokkema T, Kooiman TJM, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. *Med Sci Sports Exerc* **2017**;49:793–800.
18. Husted HM, Llewellyn TL. The accuracy of pedometers in measuring walking steps on a treadmill in college students. *Int J Exerc Sci* **2017**;10:146–153.
19. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of activity monitor step detection is related to movement patterns. *J Phys Act Health* **2016**;13:145–153.
20. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *Int J Behav Nutr Phys Act* **2008**;5:56.
21. An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? *Eur J Sport Sci* **2017**;17:360–368.

Address correspondence to:  
**Matilde Grøndahl Petersen, MSc**  
*Department of Engineering*  
*Aarhus University*  
*Finlandsgade 22*  
*Aarhus N 8200*  
*Denmark*

**E-mail:** [matilde\\_groendahl@hotmail.com](mailto:matilde_groendahl@hotmail.com)

**Received:** September 27, 2017

**Revised:** October 10, 2017

**Accepted:** October 11, 2017

**Online Publication Date:** January 4, 2018