Gait & Posture 80 (2020) 80–83

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Short Communication

Validation of different stepping counters during treadmill and over ground walking

Morten Bilde Simonsen[a,b,*], Mikkel Jacobi Thomsen[a], Rogerio Pessoto Hirata[a]

[a] Sport sciences, Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 7, DK-9220 Aalborg East, Denmark
[b] Centre for Clinical Research, North Denmark Regional Hospital, Bispensgade 37, DK-9800 Hjoerring, Denmark

A R T I C L E I N F O

Keywords:
Activity tracker
Treadmill
Step count
Validation

A B S T R A C T

Background: Commercially available physical activity trackers are very popular in the general population and are increasingly common in clinical and research settings. The marketfor activity trackers are rapidly expanding, requiring them to be validated on an ongoing basis. Different approaches have been used for validating these devices. Studies using treadmills shows good step-counting accuracy although test performed in field tests settings are limited.
Research question: Does step-counting validity differ between a field test and a treadmill protocol for different types of activity trackers?
Methods: Thirty healthy subjects participated in this study, mean age was 28.2 (± 4.33) years, body mass 78.9 (± 12.9) kg, and height 178.5 (± 9.7) cm. A treadmill protocol with three different walking speeds (2, 3 and 4 km/h) and a 982 m field test was used. During the tests, participants' feet were filmed using a waist-mounted camera. The number of steps were extracted from the video data and used for comparison with four different step counters: a) Polar M200; b) Polar A300; c) Dunlop pedometer; d) Samsung Galaxy S9 smartphone. Validity and agreement determined was determined with the use of Bland-Altman plot and Spearman's correlation.
Results: Validity was higher for the field test compared to the 4 km/h treadmill test for all tested devices. The smartphone was the most accurate in terms of error, validity and agreement for both the treadmill and field test. All devices performed poorly for the 2 km/h treadmill test and only the smartphone performed well at 3 km/h.
Significance: The results of this study show that step counting validity and error obtained during treadmill walking is not similar to a field test. Future validation studies of activity trackers should consider this when designing a protocol. The smartphone had the lowest mean bias during the field test.

1. Introduction

Commercially available physical activity trackers have grown rapidly in popularity during the last decade and are integrated in most new smartphones and smartwatches [1]. Smartphones are increasing becoming more accessible worldwide. Therefore, they have the potential to be used as a low cost indicator for activity in both clinical and research settings. Applications within research and clinical settings set very high requirements to error and reliability of these devices, with accuracy studies having found varying degree of accuracy [2–4]. However, step-counting accuracy studies have also detected limitations of activity trackers during different conditions, e.g. greater errors during slower walking speeds [4–7] and among individuals with high body mass index or gait disorders [8–10]. The majority of the studies testing step-counting accuracy have been performed in controlled laboratory environments using treadmills [1,3,4] although the laboratory conditions may not be a good proxy of real world scenarios [4,7].

The purpose of this study was to test different types of activity trackers during a treadmill and a field test protocol. *It was hypothesized better validity scores will be achieved (I) during the 4 km/h treadmill test compared to the 2 km/h, 3 km/h; and (II) during 4 km/h treadmill test than the field test.*

2. Methods

2.1. Subjects

The minimum required sample size for this study to able to detect a

* Corresponding author at: Sport sciences, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Fredrik Bajers Vej 7D-3, 9220 Aalborg E, Denmark.
E-mail address: mbsi@hst.aau.dk (M.B. Simonsen).

https://doi.org/10.1016/j.gaitpost.2020.05.037
Received 8 July 2019; Received in revised form 6 March 2020; Accepted 22 May 2020
0966-6362/

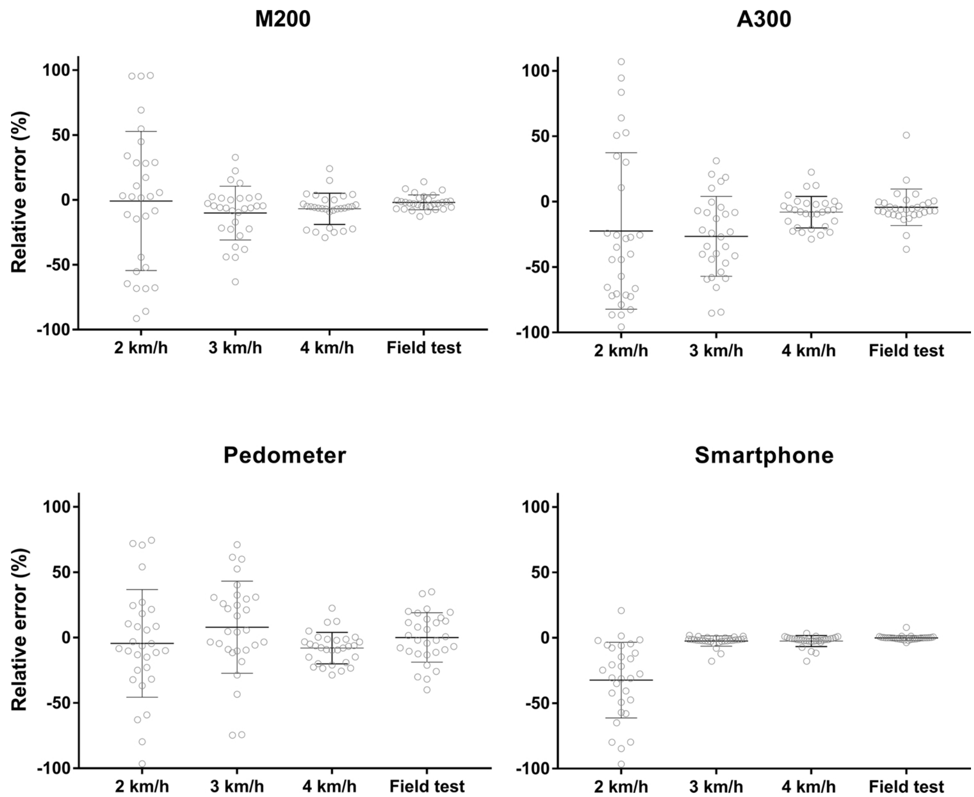

Fig. 1. Relative error scores for the treadmill (2, 3 and 4 km/h) and field test for all subjects (n = 30). The dots indicate each individuals data point and the bars are the mean ± standard deviation.

significant correlation of at least 0.5 with sufficient power (80 %) was 29 subjects [11]. Thirty healthy subjects participated in this study. All subjects had previous experience with treadmill walking. Mean age was 28.2 (± 4.33) years, body mass 78.9 (± 12.9) kg, and height 178.5 (± 9.7) cm (13.3 % females).Subjects were recruited among staff and students through adverts placed around the university.

2.2. Ethics

This study was conducted in accordance to the Declaration of Helsinki. Written and oral informed consent was obtained from all participants. The study was presented to the North Denmark Region Committee on Health Research Ethics, who stated that formal approval was not necessary for this study according to Danish laws [12].

2.3. Devices

Four commercial devices were tested in the present study: two smartwatches (Polar M200 and A300 (Polar, Kempele, Finland)), a smartphone (Samsung Galaxy S9 (Version 9) (Samsung, Seoul, South Korea) with Samsung Health (Version 6.5.0.039) installed and a pedometer (Dunlop Sport, Surrey, United Kingdom). The pedometer was mounted in a belt placed close to the right anterior superior iliac spine, the smartphone was placed in the right front pocket of the participants trousers and both smartwatches were placed distally on the left forearm. The devices were worn simultaneously. Prior to each test, the actual daily step count from the devices was noted and used for zeroing the step count values from the devices

2.4. Treadmill test

A treadmill-based protocol was employed in order to ensure that gait speeds could be standardized across all subjects. Subjects walked on the Woodway Pro XL treadmill (Woodway World, Waukesha, USA) at 3 different walking speeds (2 km/h, 3 km/h and 4 km/h), each lasting 3 min (all 3 min for each speed was used for data analyses) [7]. A short pause was provided between each gait speed. Only the 4 km/h speed was used for comparison with the field test, since it was the closest to the walking speed during the field test (3.76 ± 0.2 km/h).

2.5. Field test

Subjects walked a 982 m route around the university campus representing different everyday challenges, such as stairs, outdoors, indoors, hills and turns at a self-selected speed. A detailed description of the sectors is provided in supplementary material. Stairs walking distance was calculated by using the formula from Huang et al. (2016) [13].

2.6. Extracting of video score

Subjects wore a GoPro HERO4 Session camera ((GoPro Inc., San Mateo, USA), 30 fps and 720p) positioned in a chest mount filming the feet of the subjects. Two independent investigators each watched half of the video recordings and counted the steps (video score - VS). Inter rater reliability between two investigators have previously been reported as being reliable (ICC = 0.96) [4].

2.7. Error scores

Standard error of the estimate (SES) and mean error were calculated. Relative error was calculated using the formula:

$$\frac{\text{Device} - \text{Video score}}{\text{Video score}} \bullet 100$$

Table 1
The table is presenting bias (difference in measurement with the devices and video counted steps), 95 % limits of agreement from the Bland & Altman plot, correlation coefficients and the standard error of the estimate for all treadmill walking speeds and the field test for all devices. The "*" indicates significant bias from the video score. All P-values are Bonferroni corrected.

Test	Device	Mean bias (Steps)	Limits of agreement			Spearman's r	Validity P-value	Standard error of the estimate (Steps)
Treadmill 2 km/h	M200	5.3	−260	–	270	−0.19	0.30	20.39
	A300	−77	−390	–	240	−0.23	0.20	22.43
	Pedometer	−16	−150	–	120	0.10	0.58	22.40
	Smartphone	−62*	−190	–	70	0.52	< **0.01**	9.17
Treadmill 3 km/h	M200	−16	−110	–	81	0.29	0.11	14.15
	A300	71*	−220	–	74	0.24	0.19	14.28
	Pedometer	15	−150	–	180	0.22	0.23	14.16
	Smartphone	−5*	−11	–	0.8	0.86	< **0.01**	9.41
Treadmill 4 km/h	M200	−19*	−62	–	25	0.23	0.20	19.87
	A300	−22*	−100	–	56	0.13	0.48	19.78
	Pedometer	5.5	−120	–	130	0.32	0.07	19.36
	Smartphone	−5*	−15	–	5.2	0.76	< **0.01**	15.94
Field test	M200	−61*	−200	–	76	0.67	< **0.01**	57.01
	A300	−97*	−16	–	−180	0.49	< **0.01**	44.37
	Pedometer	13	−300	–	330	0.54	< **0.01**	94.48
	Smartphone	−5.5	−50	–	39	0.88	< **0.01**	30.25

2.8. Statistics

Data were not normally distributed and therefore non-parametric tests were used. Criterion-related validity of the devices compared to the VS was determined by a Spearman's correlation test with Bonferroni correction. Differences between devices and VS was tested with Wilcoxon signed ranks test, with Bonferroni correction. Measurement agreement between devices and the VS steps was visually inspected with non-parametric Bland & Altman plots [14] since data was skewed. All statistical analyses were performed using SPSS 23 (SPSS Inc., Chicago, IL) and Matlab R2018B (The MathWorks, Inc., Natick, Massachusetts, USA), with a significance level of 0.05.

3. Results

3.1. Effect of different treadmill walking speeds on step counting validity

Relative errors in all devices, except the pedometer, decreases when the speed increased (from 2 km/h to 4 km/h, Fig. 1). There is not a consistent trend for the mean bias for M200, A300 and Pedometer across the different speeds (Table 1). However, the smartphone mean bias consistently decreased while increasing walking speed (-5 steps for speed 3 and 4 km/h and -62 steps for 2 km/h, Table 1). A systematic reduction of limits of agreement is observed for M200 and A300 from speed 2 km/h to 4 km/h (Table 1) which can also be visually observed in the data error in Fig. 1. Similar limits of agreements are observed across all speeds for the pedometer (Table 1). Finally, the limits of agreement for the smartphone consistently decreased during speeds 3 and 4 km/h compared with 2 km/h.

A positive correlation between the VS and smartphone at 2 km/h ($r = 0.52$, SES = 9.17, p-value < 0.01). However, the Bland-Altman plots and Table 1 shows large limits of agreement and a positive trend in the bias indicating the mean difference rises with increasing number of steps (Fig. 2) [15]. Only the smartphone showed a positive correlation with VS for the speeds 3 ($r = 0.86$, SES = 9.41, p-value < 0.01) and 4 km /h ($r = 0.76$, SES = 15.94, p-value < 0.01).

3.2. Step-counting validity during the field test

During the field test, the smartphone device had the highest correlation with VS ($r = 0.88$, SES = 30.25, $p < 0.01$), followed by Polar M200 ($r = 0.67$, SES = 57.01, $p < 0.01$), Pedometer ($r = 0.54$, SES = 94.48, $p < 0.01$) and Polar A300 ($r = 0.49$, SES = 44.37, $p < 0.01$), Table 1. The smartphone had the lowest mean bias of -5.5 steps followed by the pedometer (13 steps), Polar M200 (-61 steps) and Polar A300 (-97 steps). Bland-Altman plots indicate a positive trend in the bias for pedometer, which also had the largest limits of agreement between the devices (Fig. 2).

4. Discussion

The smartphone device was the only valid device during the treadmill speeds of 3 and 4 km/h although a significant bias was observed in those conditions. In line with previous studies [4–7] all devices performed poorly during the slower gait speed of 2 km/h. The smartphone performed best during the field test with a good validity and not significant bias. In general, the other conditions presented a positive trend in the bias, indicating the number of steps taking increases the bias (up to approximately 10 % of the total number of steps for the A300 during the field test, to as little as 0.5 % for the smartphone in the same condition). This information, combined with the bias quantified in this study, may be important for improvements in future algorithms on these devices. Contrary to our hypothesis, the treadmill test was the most inaccurate in terms of error/bias and agreement for three of the tested devices.

This study only tested the devices during walk-based activities although daily activities include other types of movement patterns. Further if applied in real world scenarios participants might carry their phone in a handbag/backpack which was not considered in this study. Therefore, when selecting a device for research use, one must balance accuracy, validity and feasibility.

In conclusion, this study shows that treadmill validation scores are not identical to over-ground walking and, the smartphone device seems to be the best choice for estimation of step counting.

Declaration of Competing Interest

The authors have no real or perceived financial and personal relationships with other people or organizations that could inappropriately influence (bias) our work. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations.

Acknowledgements

This study was supported by Trygfonden (www.trygfonden.dk/english) (grant number 124714) and the Danish Rheumatism Association (https://www.gigtforeningen.dk/for-forskere/the-danish-rheumatism-association/), (grant number R161-A5276). www.

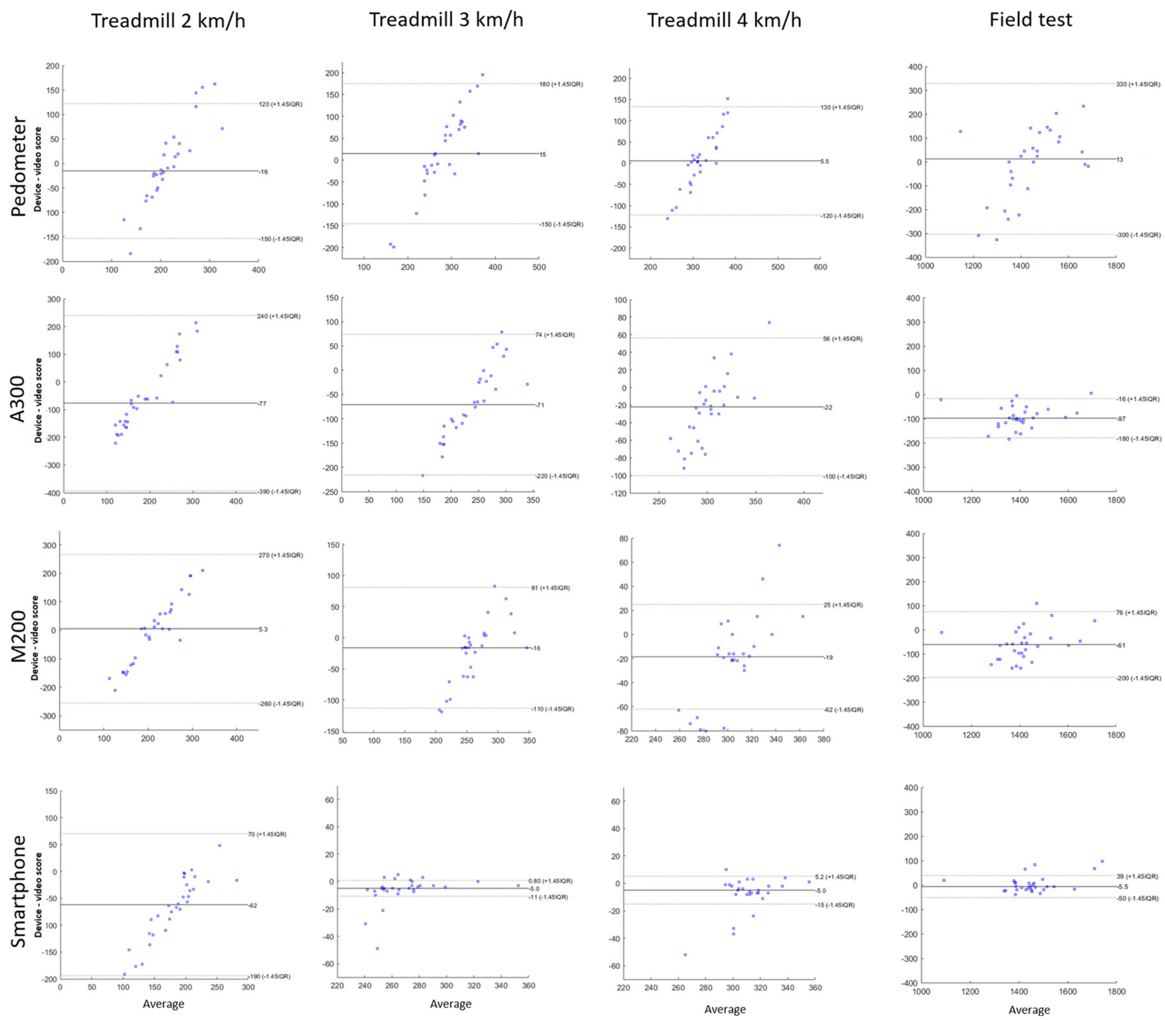

Fig. 2. Bland & Altman plot for the treadmill and field test (n = 30). Data is based on raw values for video counted steps and steps counted by each device. The solid line indicates the bias and the broken lines indicate the upper and lower limits of agreement (± 1.45 interquartile range).

trygfonden.dk/english

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.gaitpost.2020.05.037.

References

[1] A. Henriksen, M.H. Mikalsen, A.Z. Woldaregay, M. Muzny, G. Hartvigsen, L.A. Hopstock, S. Grimsgaard, Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables, J. Med. Internet Res. 20 (2018), https://doi.org/10.2196/jmir.9157.

[2] T. Fokkema, T.J.M. Kooiman, W.P. Krijnen, C.P. Van Der Schans, M. De Groot, Reliability and validity of ten consumer activity trackers depend on walking speed, Med. Sci. Sports Exerc. 49 (2017) 793–800, https://doi.org/10.1249/MSS.0000000000001146.

[3] J.J. Chow, J.M. Thom, M.A. Wewege, R.E. Ward, B.J. Parmenter, Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site, Gait Posture 57 (2017) 199–203, https://doi.org/10.1016/j.gaitpost.2017.06.012.

[4] C.K. Wong, H.M. Mentis, R. Kuber, The bit doesn't fit: evaluation of a commercial activity-tracker at slower walking speeds, Gait Posture 59 (2018) 177–181, https://doi.org/10.1016/j.gaitpost.2017.10.010.

[5] F.A. Storm, B.W. Heller, C. Mazzà, Step detection and activity recognition accuracy of seven physical activity monitors, PLoS One (2015), https://doi.org/10.1371/journal.pone.0118723.

[6] K.L. Storti, K.K. Pettee, J.S. Brach, J.B. Talkowski, C.R. Richardson, A.M. Kriska, Gait speed and step-count monitor accuracy in community-dwelling older adults, Med. Sci. Sports Exerc. (2008), https://doi.org/10.1249/mss.0b013e318158b504.

[7] C.B. Thorup, J.J. Andreasen, E.E. Sørensen, M. Grønkjær, B.I. DInesen, J. Hansen, Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease, BMJ Open 7 (2017) 15–17, https://doi.org/10.1136/bmjopen-2016-011742.

[8] J. Lauritzen, A. Muñoz, J. Luis Sevillano, A. Civit, The usefulness of activity trackers in elderly with reduced mobility: a case study, in: stud, Health Technol. Inform. (2013), https://doi.org/10.3233/978-1-61499-289-9-759.

[9] G.D. Fulk, S.A. Combs, K.A. Danks, C.D. Nirider, B. Raja, D.S. Reisman, Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury, Phys. Ther. (2014), https://doi.org/10.2522/ptj.20120525.

[10] S.A. Clemes, S. O'Connell, L.M. Rogan, P.L. Griffiths, Evaluation of a commercially available pedometer used to promote physical activity as part of a national programme, Br. J. Sports Med. (2010), https://doi.org/10.1136/bjsm.2009.061085.

[11] W.C. Guenther, Desk calculation of probabilities for the distribution of the sample correlation coefficient, Am. Stat. (1977), https://doi.org/10.1080/00031305.1977.10479195.

[12] T.D.N.C.O.H.R. ETHICS, What to notify? (2019) (accessed July 4, 2019), http://en.nvk.dk/how-to-notify/what-to-notify.

[13] Y. Huang, J. Xu, B. Yu, P.B. Shull, Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking, Gait Posture 48 (2016) 36–41, https://doi.org/10.1016/j.gaitpost.2016.04.025.

[14] R. Klein, Bland-Altman and Correlation Plot, (2019) (accessed July 1, 2019), https://www.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot.

[15] J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against standard method is misleading, Lancet. (1995), https://doi.org/10.1016/S0140-6736(95)91748-9.