IJSPT

ORIGINAL RESEARCH

VALIDITY AND RELIABILITY OF THE FITBIT FLEX™ AND ACTIGRAPH GT3X+ AT JOGGING AND RUNNING SPEEDS

Denise Jones[1]
Kay Crossley[1]
Ben Dascombe[1]
Harvi F. Hart[1]
Joanne Kemp[1]

ABSTRACT

Background: Monitoring levels of physical activity, as an outcome or in guiding rehabilitation, is challenging for clinicians. Personal activity monitors are increasing in popularity and provide potential to enhance rehabilitation protocols. However, research to support the validity and reliability of these devices at jogging and running speeds is limited.

Purpose: The purpose of this study was to evaluate the validity of the Fitbit Flex™ and ActiGraph GT3X+ for measuring step count at jogging and running speeds. A secondary purpose was to examine inter-device reliability of the Fitbit Flex™.

Study Design: Cross-sectional study

Methods: Thirty healthy participants aged between 19 and 50 years, completed a treadmill protocol at jogging and running speeds (8 km/h to 16 km/h). Treadmill speed was progressively increased by intervals of 2 km/h. Each interval was four minutes in duration with a two minute rest period between stages. Participants were encouraged to continue through the graded exercise test until they reached the maximum running speed that they felt they could maintain for four minutes. Step count data was collected for Fitbit Flex™ devices and the ActiGraph GT3X+. Video analysis of step count was used as the criterion measure.

Results: At speeds of 8 to 14 km/h Mean Absolute Percentage Errors were ≤1% for the Fitbit Flex™ and the ActiGraph GT3X+ when compared to step count via video analysis. Standard Error of Measurement between the three Fitbit Flex™ devices was ≤7 steps for speeds of 8 to 14 km/h and varied between 9 to 19 steps at 16 km/h. Fitbit Flex™ devices showed good to excellent between device reliability at speeds of 8 to14 km/h (ICC 0.723 to 0.999; p ≤0.001). Greater variability was evident with the low participant numbers at 16 km/h (ICC 0.527 to 0.896; p ≥ 0.02).

Conclusion: Both the Fitbit Flex™ and the ActiGraph GT3X+ provide a valid account of steps taken at jogging and running speeds up to 14 km/hr, attainable by non-elite runners on a treadmill. Fitbit Flex™ devices provide equivalent step count output to each other, enabling comparison between devices during treadmill jogging and running.

Level of evidence: 2b

Key words: Accelerometer, activity tracker, activity monitor, physical activity, step count

[1] La Trobe Sport and Exercise Medicine Research Centre (LASEM), School of Allied Health, College of Science, Health and Engineering, La Trobe University, Melbourne, Victoria, Australia

The authors affirm that they have no financial affiliation (including research funding) that has direct financial interest in any matter included in this manuscript.

This work is supported by an Australian Government Research Training Scholarship. Dr H. F. Hart is supported by a National Health and Medical Research Council (Australia) Project Grant (GNT1106852).

The authors would like to thank Laura Anderson for her assistance in data collection as well as the contributions of Sally Coburn, Dan Bonano, Mathew Francis, Christian Barton, Anthea Clarke, and Nirmarla Perera. We would like to thank all the participants for their time and effort in contributing to this study.

CORRESPONDING AUTHOR
Denise Jones
La Trobe Sport and Exercise Medicine Research Centre (LASEM), School of Allied Health, College of Science, Health and Engineering
La Trobe University, Melbourne, Victoria, 3086, Australia.
Telephone: +61 (0)3 9479 2169.
Fax: +61 (0)394795768.
Email: jones.d@students.latrobe.edu.au

INTRODUCTION

Physical therapists involvement in enabling and promoting physical activity is well established.[1] Enabling the maintenance, return to or improvement of physical activity levels as a key aim of therapy interventions aligns with the scope of practice descriptors identified by the World Confederation of Physical Therapy (WCPT).[2]

As physical activity is a primary factor associated with maintaining health and wellbeing, particularly when considering all-cause mortality,[3-5] it is undoubtedly an important outcome for athletes and non-athletes alike. Although injury has been shown to have a profound effect on long-term activity, irrespective on ongoing disability,[6] identifying suitable and user-friendly methods for monitoring and guiding physical activity is challenging for individuals, clinicians and researchers.

Step count is frequently used as an indicator of physical activity, the number of steps identifying a volume, rather than intensity of activity. Intensity may be extrapolated from the number of steps taken in a given time. It furnishes clinicians with a simple measure to provide guidelines and encourage behavior change for individuals and communities. This utility assumes that devices are reporting an appropriate account of steps taken. Evidence is currently lacking to substantiate the accuracy of step count output from devices in relation to more athletic populations.

Similarly, step count has been used to monitor post-intervention progress in individuals with a health condition, particularly where weight-bearing activity is a key healthcare outcome.[10-12] For runners, an accurate perception of the number of steps taken per minute may also be of relevance in relation to rehabilitation, such as attempting to increase step rate (cadence) to reduce patellofemoral load.[13] The increasing popularity of personal fitness trackers is indicative of individual enthusiasm for monitoring activity data. In addition, these trackers are serving to take the collection of objective physical activity data beyond the laboratory and into the public domain. The popularity of these devices provides opportunities for measuring physical activity that researchers and healthcare professionals are beginning to exploit. As with all emerging technologies, the purpose-specific utility of these devices needs to be established. Fitbit remains at the forefront of the market in digital fitness devices,[14] the Fitbit Flex™ being a popular wrist-worn device available at a relatively affordable price (~ USD$60).

Current research focuses on the validity of devices at lower speeds, which may be relevant for populations with chronic conditions that inhibit aerobic activity levels.[15-21] For clinicians working with sporting populations, and communities who are capable of running, these boundaries need to be expanded to evaluate the utility of devices at greater ambulation speeds. Correlation estimates of step count for the Fitbit Flex™ vary between studies. For speeds between 3 and 8 km/h, Diaz et al.[16] report strong correlations to criterion measure (0.77 to 0.85), conversely, Sushames et al.[21] report intraclass correlations of 0.05 and 0.34 for step count during walking and jogging respectively. Huang et al.[18] reported Mean Absolute Percentage Errors (MAPE's) of 6.5% and 8.9% for the Fitbit Flex™ at treadmill speeds of 3.24 and 6.41 km/h, respectively. During combined walking and jogging, Nelson et al.[22] reported a comparable MAPE of 6%. Although study protocols vary, a tendency for the Fitbit Flex™ to underestimate step count is evident, this effect being more pronounced at slower speeds.[15,16,18,21] Data is limited to substantiate the performance of the Fitbit Flex™ at speeds above 8 km/h. Therefore, this study investigated the validity of the Fitbit Flex™ at jogging and running speeds by assessing accuracy of the device output in relation to observed values for step count. Inter-device reliability was assessed by evaluating the precision of output between Fitbit Flex™ devices over the same range of jogging and running speeds. In comparison to commercially available activity trackers, the ActiGraph GT3X+ is a research grade device which allows access to underlying algorithms and options for the user in converting raw count data to step count and energy expenditure data. It is frequently used as a comparator to commercially available devices in assessing physical activity.[23-27] Simultaneous investigation of the Fitbit and ActiGraph devices was undertaken to provide comparative measures to aid assessment of their relative merits for researchers.

The purpose of this study was to evaluate the validity of the Fitbit Flex™ and ActiGraph GT3X+ for measuring step count at jogging and running speeds. A secondary purpose was to examine inter-device reliability of the Fitbit Flex™. The results of this study provide an objective measure of interest to the running community using the Fitbit Flex™ for personal activity monitoring and guidance for clinicians wishing to utilize these devices within rehabilitation and maintenance programmes, such as implementing graded return from injury or embedding modifications to running step rate to modify joint loading.

METHODS

Participants

Thirty young and middle-aged healthy adults were recruited for this cross-sectional study. Participants were recruited within the university and the wider community via website postings, social media, and word-of-mouth. The study was approved by La Trobe University Human Ethics Committee (Approval number HEC16-082).

Eligibility criteria

The Physical Activity Readiness Questionnaire (PARQ) was used to screen for safe participation. Potential participants were excluded on the basis of acute or chronic health conditions that precluded running activity, being pregnant, breastfeeding, being outside the age range of 18 to 50 years or lacking sufficient English language skills to give informed consent.

Equipment

ActiGraph

The ActiGraph GT3X+ (ActiGraph, Pensacola, FL) is a small (4.6 x 3.3 x 1.5 cm), lightweight (19g) tri-axial accelerometer (Figure 1A). It was worn on an elastic belt below the waist, in line with the right anterior axillary line and did not impede the participants' ability to run.

Fitbit

The Fitbit Flex™ (Fitbit Inc., San Francisco, CA) is a consumer-wearable activity tracker. The triaxial accelerometer is held within a wristband providing a five-light LED display of activity progress (Figure 1B).

Protocol

Data collection took place in a non-air-conditioned physiology laboratory at La Trobe University, Melbourne, Australia, between December 2016 and February 2017. Prior to undertaking testing, potential participants were offered further information on the study and screened for eligibility. All participants

Figure 1. *A. ActiGraph GT3X+ (With permission, Actigraphcorp.com); B. Fitbit FlexTM (With permission, fitbit.com)*

gave written, informed consent prior to undertaking the test.

Self-reported measures of body mass (kg) and height (m) were used where recent accurate measures could be offered by participants. Where any queries arose, measurements were confirmed in the testing laboratory using a stadiometer and digital scales. Body mass index (BMI) was calculated from these measures.

Participants were advised to wear suitable sports clothing and footwear for the test and abstain from alcohol, caffeine, and cigarettes for 24 hours prior to the test. Participants were also advised to avoid a large meal for at least three hours prior to testing and avoid vigorous exercise during the 24 hours prior to testing in line with standard recommendations for maximal exercise testing.[28]

Prior to use by each participant the ActiGraph GT3X+ devices were initialized via the supporting software (Actilife 5.10.0, ActiGraph, Pensacola, FL) inputting: start time; sampling rate (30Hz); device position; date of birth; sex; body mass; height and race of the participant.

Each participant was fitted with three Fitbit Flex™ devices, two on the left wrist (device numbers 1 and 2) and one on the right wrist (device number 3). Each band was securely fitted to the participant's wrist to allow minimal movement during testing without being uncomfortable.

For each Fitbit Flex™ device participants' demographic data (sex, date of birth, height, body mass, walking and running stride length) were entered, via the web Fitbit interface. Fitbit defines a stride as heel strike to heel strike of the opposite foot, more conventionally defined as a step. This was assessed for individual participants with a measured 10-step walk in a straight line over flat ground. The process was repeated at a comfortable running pace, self-selected by the participant. The distance was then divided by a factor of 10 to give an average 'stride'/step length.

For the purposes of treadmill testing, participants were deemed to be undertaking a standardized activity with relatively symmetrical arm movement. Due to this bilateral equivalence, all the Fitbit Flex™ devices were maintained at the default setting of 'non-dominant' throughout all data collection.

To obtain minute level data from the Fitbit Flex™ devices, each device was placed into 'activity mode' by tapping the device sharply (1 to 2 sec) until it vibrated. The activity mode was deactivated at the end of each test by repeating this procedure. This allowed for a discrete set of minute-by-minute data to be viewed via the interface.

At the start of each test session, participants were given a warm-up period of five minutes on the treadmill (Cosmed T200, Rome, Italy) to familiarize them with the equipment and ensure that all devices were comfortable and secure. Participants then undertook two further warm-up periods at 4 and 6 km/h for four minutes at each level, separated by two minutes rest, to familiarize them with the treadmill protocol. Participants were advised to maintain their regular arm swing, avoid looking at the devices and to avoid holding on to the treadmill.

The graded exercise test began at 8 km/h, progressing at 2 km/h intervals. Each interval was four minutes in duration with a rest period of two minutes between each interval. Rest periods facilitated transitions and the tracking of data between devices. Intervals were recorded for video analysis of step count. A video camera (Lumix DMC-FZ2000, Panasonic, UK) was placed to capture right and left footfall during each incremental stage of a test. A clearly visible digital clock was placed within the video frame to enable tracking of real time. Data from devices were compared to video observation of step count, which was regarded as the criterion measure.

Participants were encouraged to continue through the graded exercise test until they reached the maximum running speed that they felt they could maintain for four minutes. Each test was terminated either at the participant's request or at a point at which the researchers had concerns for the participant's wellbeing.

Data Processing

Following test sessions, each Fitbit Flex™ was synced to allow the data to be accessed via the product interface. Data from each ActiGraph GT3X+ was downloaded via a universal serial bus (USB) and processed using proprietary software (Actilife 5.10.0, Actigraph Corp. Pensacola, FL). The data were processed in 60-second epochs to align with the output from the Fitbit Flex™.

Videos were downloaded to a PC and viewed via Windows Media Player. The recordings (30 frames per second) were visually analyzed in slow motion and the number of steps, identified by foot strike, tallied for the middle two minutes of each level completed by the participant. The middle two minutes of each stage was used to minimize inconsistencies related to participants settling into their target pace or becoming fatigued at the termination of later stages The observed video data provided criterion values for step count at each level of the treadmill test. A proportion (10%) of the step count data was analyzed by two assessors (DJ and SC) to ensure consistency.

Analysis

Sample size

Sample size numbers were determined by procedures described by Walter et al.[29] for inter-device reliability. Twenty two subjects were deemed to be acceptable to judge the difference between two devices with a minimally acceptable level of 0.5, when $\alpha = 0.05$ and $\beta = 0.20$ (power = 0.08). A sample size of 18 participants was required to assess validity of the devices based on an estimated correlation coefficient of r = 0.6, 2 tailed test ($\alpha_2 = 0.05$) with a power of 80%.[30]

Inter-device reliability

Inter-device reliability was determined for the three Fitbit Flex™ devices using intraclass correlation coefficients (ICC 2, 1)[30] with 95% confidence intervals (CI). ICC's were considered to be excellent (0.75 and 1.00); good (0.60 and 0.74); fair (0.40 and 0.59) or poor ($\leq$0.40).[31] Paired *t*-tests ($\rho = 0.05$) were performed on normally distributed data to determine the mean difference (group mean difference) between devices. The standard error of measurement (SEM) was calculated for normally distributed data to determine absolute reliability. This was calculated using the formula SEM = Standard deviation (SD) x $\sqrt{1\text{-ICC}}$.

Validity

Validity was evaluated for the Fitbit Flex™ and ActiGraph GT3X+ for step count, by comparing to observed step count. Correlations between device and criterion measure were judged on the following guidelines for correlation coefficient (r): Little or no relationship (0.00 to 0.25); fair relationship (0.25 to 0.50); moderate to good relationship (0.50 to 0.75) and good to excellent relationship (above 0.75).[30]

To further investigate device validity, MAPE was used to provide a conservative estimate of individual level error.[32] MAPE is calculated with the following formula:

$$\frac{\text{Absolute bias (criterion – device)}}{\text{Criterion}}$$

Limits of agreement were used to show the spread of the difference of scores.

The significance criteria for all tests was $\alpha = 0.05$ and $\beta = 0.20$, thus power = 0.8 (1-β), and confidence intervals were 95% (1-α).

RESULTS

Participant characteristics

Between November 2016 and February 2017, 54 potential participants responded to notification and advertisement of the study. Figure 2 summarizes the flow of respondents through the study. Thirty healthy adults (18 women, 12 men; mean ± SD: age, 33 ± 8 years; BMI, 24.1 ± 2.5 kg/m^2) were included in the study (Table 1).

Findings

All 30 participants completed the protocol to the end of 8 km/h. As the speed increased above 8 km/h, there was a decrease in the sample number (Figure 3). Baseline characteristics of participants completing each level are outlined in Table 2. ActiGraph GT3X+ data were successfully obtained for all 30 participants and minute by minute data were successfully collected for all three of the Fitbit devices worn for 20 participants. For the remaining ten participants, data from two Fitbit Flex™ devices were successfully collected for seven participants. For one participant minute data was successfully collected from only one of the Fitbit devices. Two participants were missing all minute by minute data from the Fitbit Flex™ devices. The missing data was the result of errors in setting the devices to activity

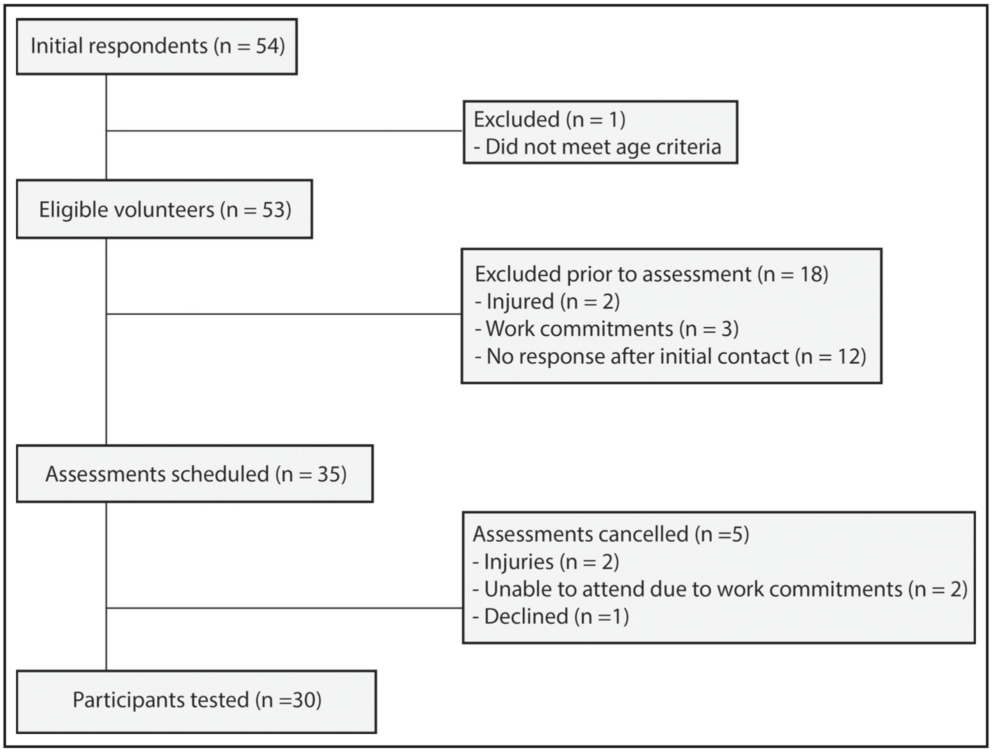

Figure 2. *Flow chart showing the number of respondents and reasons for drop-out.*

Table 1. *Demographic characteristics of the study cohort.*

Characteristics	Total Mean (SD) range n = 30	Women Mean (SD) range n = 18	Men Mean (SD) range n = 12
Age (years)	33 (8) 19 - 50	34 (7) 19 - 46	32 (8) 23 - 50
Height (m)	1.71 (0.12) 1.47 - 1.95	1.64 (0.07) 1.46 - 1.76	1.82 (0.10) 1.61 - 1.95
Body Mass (kg)	71 (16) 44 - 128	62 (8) 44 - 74	83 (16) 68 - 128
BMI (kg/m^2)	24.07 (2.51) 19.43 - 33.53	23.19 (1.77) 19.43 - 26.50	24.99 (3.13) 21.83 - 33.53

SD = +/- 1 Standard deviation; n = number of participants; m = meters; kg = kilogram; BMI = body mass index

mode. The successful functioning of this mode for the duration of the test could not conveniently be checked until the data were downloaded and viewed following completion of the trial. For seven participants, errors occurred in video records. A total of eight two-minute intervals were, therefore, missing observed step count analysis. Due to the missing data, sample size varies throughout areas of data analysis and is reported accordingly.

Observed step count, inter-rater reliability

When comparing video analysis observed step count, inter-rater reliability between both testers was excellent (ICC = 1.000, 95% CI 0.999 to 1.000).

Inter-device reliability

The three Fitbit Flex™ devices demonstrated excellent between device reliability for step count for speeds of 8 to14 km/h (Table 3), with the exception

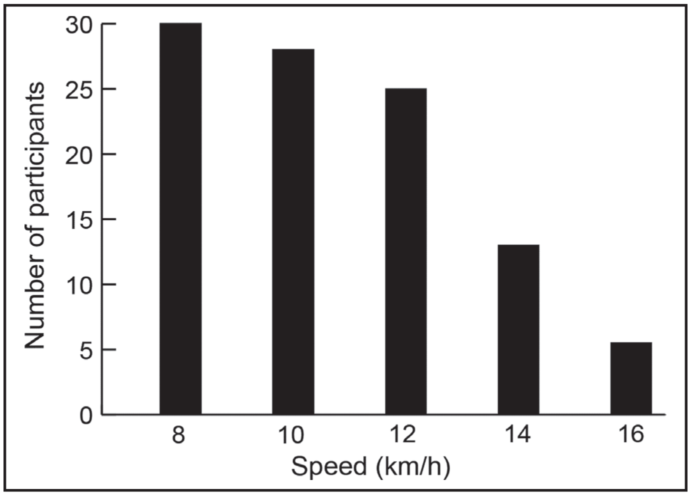

Figure 3. *Number of participants completing each level of treadmill testing.*

of Fitbit Flex™ 2 (left arm) and Fitbit Flex™ 3 (right arm) at 12 km/h, for which the intraclass correlation was good (ICC (2,1) 0.723, 95% CI 0.370 to 0.894).

The SEM between the two devices on the same arm did not vary by more than 1% at speeds of 8 to 14 km/h. This error increased to a maximum of 2% between the right and left arm devices for these speeds. Greater errors were evident at 12 km/h. A similar trend is observed at 16 km/h with SEM varying by less than 3%, at both speeds, for devices on the same side and less than 6% for devices on opposite arms.

Validity

Due to the close correlation of the output between Fitbit devices, Fitbit Flex™ 1 (left wrist) and the ActiGraph GT3X+ were assessed against the criterion measures of observed step count. Correlations between Fitbit Flex™ 1 and observed step count from video analysis were excellent for speeds of 8 to 14 km/h (Table 4). A fair relationship was evident at 16 km/h. Correlations between the ActiGraph GT3X+ and observed step count were excellent for all levels of the test, $r \geq 0.905$ for speeds of 6 to 14 km/h (Table 4). The MAPE values were <1% for both the ActiGraph GT3X+ and the Fitbit Flex™ across all reported speeds.

DISCUSSION

This study evaluated Fitbit Flex™ inter-device reliability and validity of the Fitbit Flex™ and ActiGraph GT3X+ in a healthy cohort of men and women aged 18 to 50 years. It compared the output from the Fitbit Flex™ and ActiGraph GT3X+ to the criterion measure of observed step count over speeds ranging from 8 to 16 km/h. The results indicate that both the Fitbit Flex™ and the ActiGraph GT3X+ provide a valid assessment of step count with close correlation to observed step count and MAPE values below 1% for speeds of 8 to 14 km/h.

Fitbit Flex™ inter-device reliability was excellent for devices worn on the same arm with closely associated absolute measures at speeds of 8 to 14 km/h. The low SEM between all three Fitbit devices for speeds of 8 to 14 km/h (1 to 4 steps), indicates a high level of confidence that output from the Fitbit Flex™ devices is equivalent. The large confidence intervals observed for mean differences between devices at 16 km/h highlights that participant numbers were

Table 2. *Characteristics of participants completing different levels of the test.*

Speed (km/h)	n	Men/Women	Age (Years) Mean (SD)	Height (m) Mean (SD)	Body Mass (kg) Mean (SD)	BMI (kg/m^2) Mean (SD)
8	30	12/18	34 (8)	1.71 (0.12)	71 (16)	23.80 (7.64)
10	28	12/17	33 (8)	1.72 (0.12)	71 (15)	23.95 (2.41)
12	25	11/14	33 (8)	1.72 (0.12)	70 (11)	23.66 (1.59)
14	13	8/4	34 (8)	1.77 (0.10)	75 (8)	23.83 (1.47)
16	6	5/1	34 (9)	1.80 (0.09)	77 (9)	23.72 (0.82)

km = kilometres; h = hour; n = number of participants; SD = +/- 1 Standard deviation; m = meters; kg = kilogram; BMI = body mass index

Table 3. *Fitbit Flex™ inter-device reliability for step count. Fitbit Flex™ 1 and Fitbit Flex™ 2 worn on the left wrist, Fitbit Flex™ 3 on the right wrist. Mean difference was generated from paired sample t-tests, p-value ≥0.05 indicates that output from devices does not differ significantly.*

	DEVICES	SPEED 8 KM/H	SPEED 10 KM/H	SPEED 12 KM/H	SPEED 14 KM/H	SPEED 16 KM/H
Step count Mean (SD) n	FBF 1	318 (16) 27	331 (20) 25	335 (30) 23	347 (21) 11	345 (27) 6
	FBF 2	318 (17) 21	332 (19) 19	339 (19) 17	352 (19) 8	350 (22) 4
	FBF 3	319 (15) 26	332 (19) 24	339 (30) 22	344 (20) 12	350 (18) 6
	Observed	318 (16) 27	331 (21) 26	342 (21) 24	346 (23) 10	356 (20) 5
Mean difference in step count between FBF devices (95% CI) DoF p-Value*	FBF 1 and 2	1 (0 to 2) 19 0.100	0 (0 to 1) 17 0.138	1 (-1 to 2) 15 0.542	1 (-1 to 2) 6 0.200	-3 (-23 to 18) 3 0.701
	FBF 1 and 3	1 (0 to 2) 19 0.100	1 (0 to 2) 22 0.044	-3 (-8 to 1) 20 0.137	0 (-1 to 2) 10 0.553	-5 (-29 to 19) 5 0.608
	FBF 2 and 3	0 (-1 to 0) 19 0.428	1 (-1 to 2) 17 0.284	-4 (-11 to 3) 15 0.270	1 (-1 to 2) 7 0.516	-8 (-30 to 13) 3 0.298
Standard error of measurement between FBF devices (step count)	FBF 1 and 2	1	1	4	1	9
	FBF 1 and 3	1	1	7	2	19
	FBF 2 and 3	1	2	1	2	12
ICC (95% CI) p- Value	FBF 1 and 2	0.994 (0.986 to 0.998) <0.001	0.999 (0.997 – 1.000) <0.001	0.986 (0.961 to 0.995) <0.001	0.996 (0.976 to 0.999) <0.001	0.896 (0.083 to 0.993) 0.020
	FBF 1 and 3	0.997 (0.994 to 0.999) <0.001	0.996 (0.990 to 0.998) <0.001	0.953 (0.887 to 0.981) <0.001	0.995 (0.983 to 0.999) <0.001	0.527 (-0.378 to 0.917) 0.112
	FBF 2 and 3	0.998 (0.994 to 0.999) <0.001	0.994 (0.985 to 0.998) <0.001	0.723 (0.370 to 0.894) <0.001	0.994 (0.971 to 0.999) <0.001	0.721 (-0.429 to 0.979) 0.085

SD = +/- 1Standard deviation; CI confidence interval; DoF – degrees of freedom**;** ***(2-tailed sig.)**;** ICC = Intraclass Correlation Coefficient; n = number of participants; km/h = kilometres per hour; FBF 1, FBF 2, FBF 3= Fitbit Flex™ devices 1,2 and 3; Statistically significant at p < 0.05.

Table 4. *Validity of devices to criterion measure. Correlation of Fitbit Flex™ 1 and ActiGraph GT3X+ to observed step count and percent agreement to observed step count for Fitbit Flex™ and ActiGraph GT3X+, calculated as Mean Absolute Percentage Error (MAPE).*

SPEED	OBSERVED STEP COUNT		FITBIT FLEX™							ACTIGRAPH GT3X+						
km/h	*Mean±SD*	*n*	*Mean±SD*	*n*	*r (p-value)**	*n*	*MAPE (%)*	*ULOA (%)*	*LLOA (%)*	*Mean±SD*	*n*	*r (p-value)**	*n*	*MAPE (%)*	*ULOA (%)*	*LLOA (%)*
8	318±16	27	318±16	27	0.997 (<0.001)[1]	25	0	1	0	318±15	29	0.997 (<0.001)[1]	27	0	1	0
10	331±21	26	331±20	25	0.994 (<0.001)[1]	24	1	1	-1	331±20	27	0.998 (<0.001)[1]	26	0	1	0
12	342±21	24	335±30	23	0.829 (<0.001)[1]	22	1	7	-5	338±28	25	0.990 (<0.001)[1]	24	1	2	-1
14	346±23	10	347±21	11	0.999 (<0.001)[1]	9	0	0	1	341±19	13	0.905 (<0.001)[1]	10	1	6	-4
16	356±20	5	345±27	6	0.409 (0.494)	5	4	19	-11	350±19	6	0.762 (0.134)	5	2	9	-6

* correlation to observed step count; *km/h* = kilometres per hour *SD* = +/- 1Standard deviation; *n* = number of participants; *r* = Pearson's r for parametric data; *MAPE* = Mean Absolute Percentage Error; *ULOA* = upper limit of agreement; *LLOA* = lower limit of agreement; [1]Statistically significant at p < 0.05

insufficient to draw conclusions regarding reliability of the Fitbit Flex™ devices at this speed. The relatively symmetrical upper limb activity expected with treadmill walking and running was reflected in the similarity of mean differences between devices on opposite sides. Greater variances evident in right/left data at 12 km/h reflect one outlying set of data. With this participant omitted from analysis, ICC's for Fitbit 1 and 3 improve from 0.953 to 0.995 ($p < 0.001$) and Fitbit 2 and 3 from 0.723 to 0.981 ($p < 0.001$).

In previous studies of the Fitbit Flex™, MAPE's have varied. Both Diaz et al.[16] and Sushames et al.[21] reported a trend of improvement as assessed treadmill speeds increased. Diaz et al.[16] reported a MAPE of 16% at 3 km/h improving to 1.8% at 8.4 km/h. Sushames et al.[21] observed self-selected walking speeds (between 5 and 6.5 km/h) and jogging speeds (between 8 and 10 km/h) with MAPE decreasing from 14.7% to 2.5% at higher speeds. Conversely, Huang et al.[18] reported an increase of 2.4% in MAPE's between 3.2 and 6.4 km/h. Findings in the current study are reflective of these previously reported figures at 8 to 10 km/h, additionally, the current study highlights that low MAPE's are also associated with speeds above those previously reported. The excellent correlations between Fitbit Flex™ 1 and observed step count from video analysis indicate a valid measure. With only five participants, it is inappropriate to draw conclusions regarding the relationship between the devices and criterion measure at 16 km/h.

Despite the ability to use filters to accommodate for slow speeds, studies of the ActiGraph GT3X+ mirror the trends seen in Fitbit Flex™ data with poor correlation to step count criterion measures at slow speeds, improving as more standard walking speeds are reached.[33,34] However, Tudor-Locke et al.[35] concluded that steps estimated by a waist-worn ActiGraph GT3X+ were not significantly different from observed step count in speeds ranging from 0.84 km/h to 11.28 km/h. The current study expands the pool of data available for the ActiGraph GT3X+, including previously unreported running speeds above 11 km/h. Correlations to observed step count for jogging and running speeds the current study, ranging from 0.905 to 0.990 ($p < 0.001$), reflect those reported by Lee et al.[34] for average walking speeds. These correlations are markedly different to those reported by Sushames et al.[21] for jogging (0.46, $p = 0.005$). Differences in methodology, such as the self-selection of jogging speed and the 6-minute duration of data collection may account for some of the differences observed.

For research purposes, commercial devices such as the Fitbit Flex™ potentially have substantial advantages in relation to cost; subjective perceptions of the device, such as being more agreeable to wear, and therefore compliance from participants. Their utility may be compromised by their commercial nature and the speed of change in the market. The Fitbit Flex™ assessed in this study had now been superseded by the Fitbit Flex2™. For researchers, this means that the pool of evidence underpinning data collection will remain limited for specific devices and that restricted information sharing from commercial producers will prevent researchers being able to give a full account of algorithms and accuracy when reporting their findings. Small-scale studies such as this can provide a compromise to mitigate some of the uncertainty of using commercial devices. Changes in the commercial market have less impact on clinical utility of devices which maintain the advantages of being accessible, affordable and broadly equivalent to previous incarnations of the same device in relation to the accuracy of basic algorithms such as step count.

LIMITATIONS

There are a number of limitations in this study that should be acknowledged. First the convenience sample of participants for this study encompassed a range of athletic abilities across a young and middle-aged cohort of healthy adults. The non-elite nature of the runners participating limited the number able to sustain speeds above 12 km/h. Lower participant numbers at 14 and 16 km/h compromises the validity of the findings at these speeds. A larger pool of participants would reduce the effect of outlying data such as that identified at 12 km/h. Additionally, utilizing laboratory-based measures of height and body mass for all participants would be recommended for future studies to eliminate the possibility of any inaccuracies, particularly in relation to more elite sporting populations. Second, minute by minute data for step count cannot be accessed via the Fitbit user interface unless the device has been put into an activity mode. This resulted in the loss of some data where the activity mode failed to activate or was inadvertently deactivated during the running trials. Third, the two-minute intervals reported provide a limited snap-shot of activity related to controlled treadmill running conditions. The results should be interpreted with caution as they cannot be extrapolated to be indicative of the performance of the devices over the range of running surfaces and physical activity occurring in free-living. Future research in less restrained conditions, using runner specific populations, would be a valuable addition to the current knowledge base.

CONCLUSIONS

Both the Fitbit Flex™ and the ActiGraph GT3X+ provide a valid account of steps taken at jogging and running speeds attainable by non-elite runners on a treadmill. Inter-device reliability for step count at jogging and running speeds indicates that individual users the Fitbit Flex™ can compare outputs between each other's devices for these activities with relative confidence. Users of these devices should be advised to wear the device on the same arm to provide the most reliable comparison of day-to-day data.

REFERENCES

1. Kunstler BE, Cook JL, Freene N, et al. Physiotherapist-led physical activity interventions are efficacious at increasing physical activity levels: A systematic review and meta-analysis. *Clin J Sport Med.* 2018;28(3):304-315.
2. World Confederation for Physical Therapy. Policy statement: Description of physical therapy. 2015; https://www.wcpt.org/policy/ps-descriptionPT. Accessed 01.05.2018.
3. Gaesser GA, Tucker WJ, Jarrett CL, et al. Fitness versus fatness: which influences health and mortality risk the most? *Curr Sports Med Rep.* 2015;14(4):327-332.
4. Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA.* 2009;301(19):2024-2035.
5. Lee I, Shiroma E, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet.* 2012;380(9838):219-229.
6. Andrew N, Wolfe R, Cameron P, et al. The impact of sport and active recreation injuries on physical activity levels at 12 months post-injury. *Scand J Med Sci Sports.* 2014;24(2):377-385.
7. Jiang L, Li Z, Lin J. The objective evaluation of daily physical activity for knee osteoarthritis patients after total knee arthroplasty. A prospective clinical trial using accelerometry. *Osteoarthritis Cartilage.* 2015;23(suppl2):A368.
8. LaLonde KB, MacNeill BR, Eversole LW, et al. Increasing physical activity in young adults with autism spectrum disorders. *Res Autism Spectr Disord.* 2014;8(12):1679-1684.
9. Tudor-Locke C, Bassett DR. How many steps/day are enough? *Sports Med.* 2004;34(1):1-8.
10. Babineau O, Carver T, Reid R, et al. Objectively monitored physical activity and sitting time in bariatric patients pre-and post-surgery. *J Obes Bariatrics.* 2015;2(2):1-5.
11. Bell DR, Pfeiffer KA, Cadmus-Bertram LA, et al. Objectively measured physical activity in patients after anterior cruciate ligament reconstruction. *J Sports Med.* 2017;45(8):1893-1900.
12. Crizer MP, Kazarian GS, Fleischman AN, et al. Stepping toward objective outcomes: a prospective analysis of step count after total joint arthroplasty. *J Arthroplasty.* 2017;32:S162-S165
13. Lenhart RL, Thelen DG, Wille CM, et al. Increasing running step rate reduces patellofemoral joint forces. *Med Sci Sports Exerc*. 2014;46(3):557.
14. Statista. Number of Fitbit devices sold worldwide from 2010 to 2016 (in 1,000s). 2016; https://www.statista.com/statistics/472591/fitbit-devices-sold/. Accessed 3.5.17.
15. Case MA, Burwick HA, Volpp KG, et al. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA.* 2015;313(6):625-626.
16. Diaz KM, Krupka DJ, Chang MJ, et al. Fitbit: An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol.* 2015;185:138-140.
17. Gamez JL, Gonzalez J, Leyva P, et al. Accuracy of fitbit activity trackers during walking in a controlled setting. *Int J Exerc Sci.* 2017;2(9):1344.
18. Huang Y, Xu J, Yu B, et al. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait Posture.* 2016;48:36-41.
19. Kooiman TJ, Dontje ML, Sprenger SR, et al. Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil.* 2015;7:24-35.
20. Montes J, Navalta J. Energy expenditure and step count analysis of the Fitbit Flex activity tracker: 1343 Board# 18 June 19. *Med Sci Sports Ex.* 2017;49(5S):362.
21. Sushames A, Edwards A, Thompson F, et al. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PloS one.* 2016;11(9):e0161224.
22. Nelson MB, Kaminsky LA, Dickin DC, et al. Validity of consumer-based physical activity monitors for specific activity types. *Med Sci Sports Exerc.* 2016;48(8):1619-1628.
23. Alharbi M, Bauman A, Neubeck L, et al. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. *Eur J Prev Cardiol.* 2016;23(24):1476-1485.
24. Chu AH, Ng SH, Paknezhad M, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PloS one.* 2017;12(2):e0172535.

25. Dominick GM, Winfree KN, Pohlig RT, et al. Physical activity assessment between consumer-and research-grade accelerometers: a comparative study in free-living conditions. *JMIR Mhealth Uhealth.* 2016;4(3):e110.

26. Gomersall SR, Ng N, Burton NW, et al. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and Actigraph accelerometry. *JMIR Mhealth Uhealth.* 2016;18(9):e239.

27. Reid RE, Insogna JA, Carver TE, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *J Sci Med Sport.* 2017;20(6):578-582.

28. Tanner R, Gore C. *Australian Institute of Sport. Physiological tests for elite athletes 2nd ed* Champaign: Human Kinetics. 2013.

29. Walter S, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17(1):101-110.

30. Portney LG, Watkins MP. *Foundations of clinical research: Applications to practice. 3 ed*: F A Davis Company, Philadelphia; 2015.

31. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6(4):284-290.

32. Bai Y, Welk GJ, Nam YH, et al. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc.* 2016;48(1):151-158.

33. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. *Med Sci Sports Exer.* 2012;44(4):733-741.

34. Lee JA, Williams SM, Brown DD, et al. Concurrent validation of the Actigraph GT3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *J Sports Sci.* 2015;33(10):991-1000.

35. Tudor-Locke C, Barreira TV, Schuna Jr JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sci Sports Exer.* 2015;47(4):839-842.