Original Research Paper

Validation of a consumer-grade activity monitor for continuous daily activity monitoring in individuals with multiple sclerosis

Valerie J Block ⓘ , Chao Zhao, Jill A Hollenbach, Jeffrey E Olgin, Gregory M Marcus, Mark J Pletcher, Roland Henry, Jeffrey M Gelfand and Bruce AC Cree

Abstract

Background: Technological advancements of remote-monitoring used in clinical-care and research require validation of model updates.

Objectives: To compare the output of a newer consumer-grade accelerometer to a previous model in people with multiple sclerosis (MS) and to the ActiGraph, a waist-worn device widely used in MS research.

Methods: Thirty-one individuals with MS participated in a 7-day validation by the Fitbit Flex (Flex), Fitbit Flex-2 (Flex2) and ActiGraph GT3X. Primary outcome was step count. Valid epochs of 5-min block increments, where there was overlap of ≥ 1 step/min for both devices were compared and summed to give a *daily total* for analysis.

Results: Bland–Altman plots showed no systematic difference between the Flex and Flex2; mean step-count difference of 25 more steps-per-day more recorded by Flex2 (95% confidence intervals (CI) = 2, 48; $p = 0.04$),interclass correlation coefficient (ICC) = 1.00. Compared to the ActiGraph, Flex2 (and Flex) tended to record more steps (808 steps-per-day more than the ActiGraph (95% CI= –2380, 765; $p < 0.01$), although the ICC was high (0.98) indicating that the devices were likely measuring the same kind of activity.

Conclusions: Steps from Flex and Flex2 can be used interchangeably. Differences in total step count between ActiGraph and Flex devices can make cross-device comparisons of numerical step-counts challenging particularly for faster walkers.

Keywords: Physical activity, motor activity, accelerometry, multiple sclerosis, validation studies as topic (mesh heading)

Introduction

Ambulatory disability is a frequent and often devastating consequence of multiple sclerosis (MS).[1,2] People with MS are more sedentary than the general population[3–7] which can amplify already disabling symptoms and increase the risk of sedentarism-related morbidity.[8,9] Remote activity monitoring can be useful for documenting activity status for clinical care by recording the amount of activity a person actually performs in daily life rather than what a person is capable of doing in a clinic evaluation.[10,11]

ActiGraph accelerometers are widely used in MS research, accounting for 75% of 61 articles reporting on ≥ 24 hours and <1 month of continuous physical activity monitoring in MS.[12,13] The waist-worn ActiGraph may be impractical for longer-term continuous use (i.e. year-long) due to inconvenience, expense and technical barriers. Ankle-worn devices such as the StepWatch have similar limitations.

Correspondence to:
Bruce AC Cree,
Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, 675 Nelson Rising Lane, Box 3206. San Francisco, CA, 94158, USA.
Bruce.Cree@ucsf.edu

Valerie J Block,
Chao Zhao,
Jill A Hollenbach,
Weill Institute for Neurosciences, Department

of Neurology, University of California San Francisco, USA

Jeffrey E Olgin,
Gregory M Marcus,
Department of Epidemiology and Biostatistics, University of California San Francisco, USA

Mark J Pletcher,
Department of Medicine, University of California San Francisco, USA

Roland Henry,
Weill Institute for Neurosciences, University of California San Francisco, USA
Department of Radiology, University of California San Francisco, USA

Jeffrey M Gelfand,
Bruce AC Cree,
Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, USA

Unobtrusive, commercially available, wrist-worn devices could be an alternative for continuous, longer-term monitoring of physical activity. In healthy adults, many such wearable devices have been validated against manual or video recording during treadmill training, and at home (over *c.* 7 days) as accurate measures of step count.[14–18]

Manufacturers often do not share the hardware or algorithm updates of the frequently changing wearable devices with the consumer (or researcher), nor is validation data readily available. Hence, in longitudinal studies where device loss and replacement are inevitable, it is important to ensure that newer generation devices have output metrics comparable to the original device used at study initiation.

For example, the Fitbit Flex 2 (Flex2) replaced the Fitbit Flex (Flex), a widely used wrist-worn accelerometer. The Flex2 is waterproof (reducing risk of not being re-donned post-ablution and thereby could increase adherence) and records swimming – an important option for exercise in some individuals with MS due to pervasive heat sensitivity.[19–21] In this study, we sought to (a) compare the output from the Flex and ActiGraph GT3X against manual step-counting [22] over 2 min in the clinic, and (b) compare the outputs from the Flex2 to that of the Flex, and ActiGraph at home over 7 days. The ActiGraph was selected as a reference because of its widespread use in MS studies.

Material and methods

Participants

Individuals enrolled in the FITriMS cohort study (Fitbit remote monitoring in MS)[22,23] were assessed at study entry (including a 2-min walk test (2MWT)). Participants were invited to participate in a 7-day home validation of Flex against an ActiGraph GT3X (Group 1) once they had completed >3 months,[24] and the Flex2 after >10 months in FITriMS (Group 2). Individuals with relapsing–remitting or progressive forms of MS from the UCSF MS Center clinic were prospectively recruited between July 2015 and April 2016. Inclusion criteria were: (a) a definite diagnosis of MS, defined by 2010 International Panel criteria;[25] (b) ≥18 years of age; (c) ability to walk for at least 2 min with or without an assistive device; (d) relapse free for ≥30 days prior to study entry; (e) Wi-Fi Internet access and the ability to follow instructions for technology maintenance. Participants were excluded if they had major musculoskeletal (e.g. recent broken

leg), cardiovascular or respiratory comorbidities (e.g. congestive heart failure) that could interfere with physical activity and potentially bias or obscure results. The definition of the 2014 Advisory Committee on Clinical Trials in MS Committee was employed for the characterization of relapsing and progressive MS phenotypes.[26] The UCSF IRB approved the study protocol. All participants provided written informed consent.

Study procedures

Neurologists evaluated disability using the Expanded Disability Status Scale, EDSS.[27]

Step count measurement techniques included:

- *Manual step counting:* research personnel recorded actual steps the participant took over a 2MWT.
- *ActiGraph GT3X (Manufacturing Technology, Inc., FL, USA):* this contains a triaxial accelerometer, is worn on an elastic band around the waist and is validated as a measure of walking behaviour in individuals with MS.[28,29] For this study the accelerometer signals were set at 30 Hz and 10-second epochs. The data were retrieved from the ActiGraph via a direct USB 2.0 connection and downloaded using ActiLife v6.13.2 software. A low frequency extension filter was applied, as per the manufacturer's guideline.[30]
- *Fitbit Flex (Flex) (Fitbit, CA, USA):* Flex contains a tri-axial accelerometer within a discreet bracelet worn around the wrist.[19] The Flex records step count as the main outcome and was previously validated in MS.[22] Commercial production of the Fitbit Flex was discontinued, and this model is being replaced with the Fitbit Flex 2.
- *Fitbit Flex 2 (Flex2) (Fitbit, CA, USA):* Flex2 contains a tri-axial accelerometer within a discreet bracelet around the wrist. The main outcome metric is steps per epoch (i.e. minute or day). The device can be worn at all times (waterproof up to 100 m).

Both the Flex and Flex2 synchronize with either smartphone applications (via Bluetooth) or a computer application (via USB dongle). The research team purchased the devices without financial support from device manufacturers.

At study entry, step count was compared between manual, Flex and ActiGraph over a 2MWT. For the 7-day home comparison in the two groups,

eligible participants were mailed a package containing: (a) a fully charged and activated ActiGraph accelerometer (set to record for 7 ± 3 days); (b) a Flex2 (only Group 2) device pre-set up with an anonymous Gmail account different from their original Flex account to ensure data quality; (c) written instructions; and (d) a pre-paid envelope to return the ActiGraph and Flex2 at the end of the monitoring period. An electronic message was sent with detailed instructions and a reminder to continue wearing their Flex simultaneously. Participants were instructed not to change their normal daily activities and to wear the Flex2 on the same wrist as the Flex (non-dominant). The ActiGraph was worn on an elastic belt around the waist over their non-dominant hip.

The Health eHeart Study clinical research platform (https://www.health-eheartstudy.org/) was used for electronic consent and remote Fitbit data collection. Participants gave authorization to link their Fitbit accounts for the study to recover data for research purposes. The Health eHeart database collected and stored the data through an application-programming interface.

Statistical analysis

Wear-time validation for the ActiGraph was generated using 60-second epochs.[31] The Troiano (2007) wear-time validation parameters were used.[32,33] At least 10 hours of valid wear time per day and 3 days of valid data were required for comparative assessment.[34] Interclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) were calculated.

Absolute step count was compared between devices over the 2MWT. To calculate steps per day, at home, valid measurements were defined as continuous 5 minute increments, or 'blocks', when there was overlap of at least 1 step/min for the two devices being compared. These blocks were then summed to give daily totals (Supplemental Figure S1). For this analysis, we compared blocks rather than total daily averages, since there were times where one device was used for higher intensity physical activity (running, swimming) while the other was not worn, even on days that were deemed valid. With this analytical method, the intervals compared included only times when both devices were being worn simultaneously.

We compared step counts among the measurement devices using the Bland–Altman approach[35] and paired *t*-test to determine the mean difference. The modified limits of agreement method was applied on the repeated measurements to compare the repeatability of the two raters.[35] Modified Bland–Altman (with correction) was used when there was a linear relationship between the mean difference of the step counts recorded by the devices and the average of the step counts from the devices being compared.[36,37] Paired two-sample *t*-tests assuming equal variance were used to determine if potential covariates (sex or disease subtype) were factors in the difference between step count measures. The 'leave-one-out' method[38] was used to assess for differential bias across the disability spectrum. The model was then trained on each EDSS group subset to determine the relationship between the step-count difference and the average step count on the two devices. A linear regression model was performed to examine the effect of the EDSS groups on the rate of step count difference change over the average of step count (slope).

R software calculated statistics and generated figures;[39] *p*-value < 0.05 was considered significant.

Results

Sixty-one participants were included in comparisons between Flex and ActiGraph to manual step counting (the gold standard for this study) during the 2MWT (Table 1).[22,24] Bland–Altman plots showed no systematic bias between devices. Compared to manual counts, the ActiGraph tended towards under-record steps whereas the Fitbit over counted compared to ActiGraph for lower daily steps (Supplemental Figure S2). There was no systematic bias between Fitbit and manual counts. Over 2MWT, ICCs were moderate–high between ActiGraph and Flex (0.59, 95% CI = 0.40–0.73. Mean difference = –6.03), Flex and manual (0.69, 95% CI = 0.53–0.80. Mean difference = 9.75) and ActiGraph and manual (0.76, 95% CI = 0.63–0.85. Mean difference = 15.79). Covariates (sex, disease subtype and age group) did not affect the results.

Thirty-six individuals participated in the home comparison of the Flex2 and ActiGraph over 7 days. Five participants returned the ActiGraph with data for different days than was recorded for the Flex2 and were excluded. No participants were excluded due to cardiovascular or musculoskeletal comorbidities. Median EDSS at study entry was 4.0 (Interquartile range, IQR 2.5–6.0); 17 (58%) were women; and 16 (52%) had a diagnosis of relapsing MS (Table 2). Sample size for each comparison, after quality control and performing the 'block' analysis for the daily

Table 1. Group 1: demographic data for step count comparison over a clinic-based 2-Minute Walk Test (2MWT) (Manual, Fitbit Flex and ActiGraph), and 7-day home-environment step count comparisons (Fitbit Flex and ActiGraph). (Adapted from Block 2016.[24])

	2MWT	Home comparison
Sample size (*n*)	61	20
Female (*n* (%))	44 (72)	14 (70)
Age in years (mean (SD))	50 (14.4)	54 (11.4)
EDSS (mean (range))	4.0 (0.0–6.5)	4.0 (0.0–6.5)
EDSS $\geq$ 4 (*n* (%))	34 (55.7)	12 (60)
Relapsing MS (*n* (%))	42 (69)	10 (50)

SD: standard deviation; EDSS: Kurtzke Expanded Disability Status Scale.

Table 2. Group 2: demographic data for 7-day home-environment step count comparisons (Flex, Flex2 and ActiGraph).

Demographics	All	ActiGraph v. Flex2	Flex2 v. Flex	ActiGraph v. Flex
Sample size: (*n*)				
All participants	31	22	18	20
Progressive MS	15	11	10	11
Relapsing MS	16	11	8	9
Female (*n* (% of MS subgroup))				
All participants	17 (58.1)	12 (54.5)	9 (50.0)	9 (45.0)
Progressive MS	6 (33.3)	5 (45.5)	4 (40.0)	3 (27.3)
Relapsing MS	12 (66.6)	7 (63.6)	5 (62.5)	6 (66.7)
p-value	0.9779	0.669	0.635	0.190
Average age (years (SD))				
All participants	53.4 (11.7)	53.4 (13.1)	54.4 (10.5)	55.2 (10.6)
Progressive MS	59.4 (7.5)	59.8 (8.3)	58.3 (6.6)	60.0 (6.5)
Relapsing MS	46.9 (12.0)	46.9 (14.1)	49.5 (12.8)	49.3 (12.0)
p-value	<0.001	0.009	0.059	0.011
EDSS at baseline (median (range))				
All participants	4.0 (0.0–6.5)	4.0 (0.0, 6.5)	4.0 (0.0, 6.5)	4.0 (0.0, 6.5)
Progressive MS	6.0 (3.5–6.5)	6.0 (3.5, 6.5)	6.0 (3.5, 6.5)	6.0 (3.5, 6.5)
Relapsing MS	2.0 (0.0–6.0)	2.0 (0.0, 5.0)	1.8 (0.0, 4.0)	2.0 (0.0, 4.0)
p-value[a]	<.001	<0.001	<0.001	<0.001
Disease duration (median (IQR))				
All participants	17.5 (8.5–23.0)	11.5 (5.2, 21.0)	11.5 (5.2, 21.0)	13.0 (5.8, 23.2)
Progressive MS	19.5 (9.0–25.0)	17.0 (6.5, 28.5)	19.5 (6.8, 31.2)	17.0 (7.5, 29.5)
Relapsing MS	13.0 (7.5–22.0)	10.0 (4.5, 13.5)	9.5 (4.8, 13.2)	10.0 (5.0, 14.0)
p-value	0.232	0.07	0.056	0.097
Daily step count: (mean (SD))		ActiGraph/Flex2	Flex2/Flex	ActiGraph/Flex
All participants	N/A	2148 (2292)/2880 (2581)	3089 (2719)/3060 (2691)	2125 (2278)/2908 (2531)
Progressive MS	N/A	1396 (2119)/1800 (2275)	2016 (2616)/1999 (2603)	1501 (2189)/2066 (2308)
Relapsing MS	N/A	2900 (2302)/3879 (2573)	4431 (2337)/4387 (2292)	2889 (2269)/3936 (2528)
p-value		0.111/0.054	0.042/0.042	0.165/0.084

MS: multiple sclerosis; SD: standard deviation: EDSS: Expanded Disability Status Scale; IQR: interquartile range.
[a]*p*-value is the difference in demographics between participants with progressive MS and relapsing MS.

step count totals were: 22 for the ActiGraph/Flex2, 18 for Flex/Flex2 and 20 for the ActiGraph and Flex. Including unique (different) individuals from Group 1 increased the sample to 33. Subjects reported that missing data for Flex and Flex2 occurred due to difficulties remembering to sync both devices on separate accounts. Missing ActiGraph data was due to forgetting or not wanting to don the device.

Flex2 averaged 808 more steps-per-day than the ActiGraph (Bland–Altman: 95% CI = 670, 946; $p < 0.01$). However, when we used modified
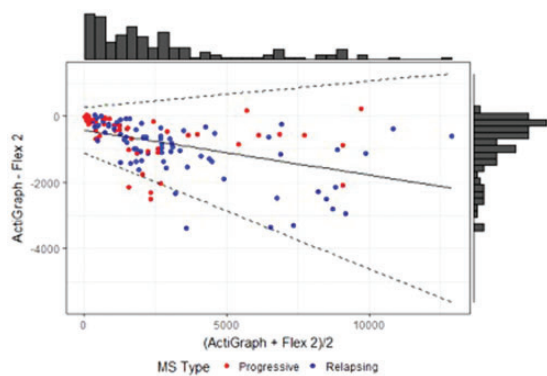
Figure 1a. ActiGraph and Flex2 (modified Bland–Altman plot).

Figure 1b. ActiGraph and Flex (modified Bland–Altman plot). The solid lines represent the mean differences between the measures and the dashed horizontal lines represent mean differences ±2 standard deviations. Each solid point corresponds to a separate individual, for a separate day. The solid lighter grey lines indicated where '0' difference between the measures would lie. The grey dashed lines are the 95% limits of agreement. The histograms on the top and right depict the distribution of the data.

(corrected) Bland–Altman there was a high percentage of data points (96%) within the 95% limits of agreement, indicating a better fit (Figure 1a, Supplementary Table 1B). A similar trend was seen between Flex and ActiGraph (mean difference: 873 per day. 95% CI = 724, 1023; $p < 0.01$) (Figure 1b). The percentage of data points within the 95% limits of agreement increased from 93 to 96% using the modified Bland–Altman, also denoting a better fit with this method (Figure 1b). No systematic difference was seen in the modified Bland–Altman plots for the Flex v. Flex2 (mean difference: 25 steps/d more recorded by the Flex2 (95% CI = 2, 48; $p = 0.04$)) (Figure 2). The percentage of data points within the 95% limits of agreement decreased (95 to 92%) when using the modified Bland–Altman, revealing that this correction did not produce a better fit.

ICCs were high for all comparisons: ActiGraph v. Flex2 (0.98 (95% CI = 0.97–0.99)), ActiGraph v. Flex (0.98 (95% CI = 0.97–0.98)) and Flex v. Flex 2 (1.0 (95%CI= 1.0–1.0)) (Table 3). A larger discrepancy was observed between devices with greater daily steps (Figures 1 and 2) and lower walking disability. The difference between ActiGraph and Flex2/Flex, was greater for individuals who took more steps/d (mean difference −387 (95% CI = −491, −282) below the cohort median v. −1229 (95% CI = −1439, −1019) above it) (Supplementary Table 2). Discrepancy was negligible between Flex and Flex2 (−10 steps difference (95% CI = −23, 4) below the cohort median, and 59 step difference (95% CI = 16, −102) above).

Between the Flex and ActiGraph, the effect of disability (increased EDSS score) on the rate of step-count difference change over the slope was not significant, although trended towards decreasing
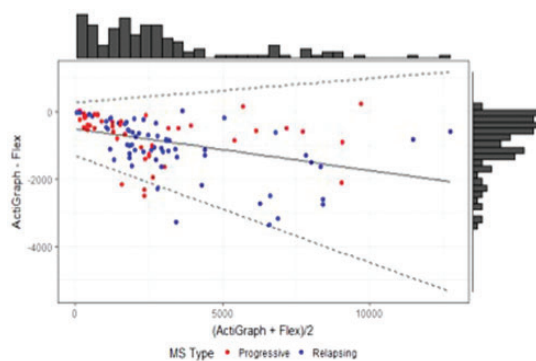
($n = 33$, β = –0.019, $p = 0.229$) (Supplementary Figure S3).

The discordance in steps recorded was greater for all comparisons in individuals with relapsing MS (Table 4). The Flex and Flex2 showed the lowest discordance in step count for both disease subtypes (mean difference progressive: 3, $p = 0.72$ (95% CI = −14, 19) and relapsing: 49, $p = 0.034$ (95% CI = 4, 94)). Between genders, significant difference in step count was observed between Flex2 and Flex for men (mean difference: 56 steps/day, 95% CI = 29, 82; $p < 0.01$) but not women (mean difference: 3 steps/day, 95% CI = −31, 38; $p = 0.848$). ActiGraph measured fewer steps than Flex2 in both women (mean difference: −830 steps/day, 95% CI = −1022, −638; $p < 0.01$) and men (mean difference: −783 steps/d, 95% CI = −988, −578; $p < 0.01$). Modified Bland–Altman plots presented a superior fit, improving the percentage of people within the 95% limits of agreement in women from 93 to 99%. Similar results were found for ActiGraph and Flex comparisons.

Discrepancy between Flex and Flex2 was greater in individuals with lower disability (EDSS < 4.0, mean difference: 49 steps/day, 95% CI = 4, 94; $p = 0.034$) than those with greater disability (EDSS ≥4.0, mean difference: 3 steps/d, 95% CI = −14 steps/day 19; $p = 0.721$). Greater discrepancy was seen between ActiGraph and Flex2 for individuals with lower disability (EDSS < 4.0: mean difference: −1086 steps/d,

95% CI –1303, –868; $p < 0.01$) than higher disability scores (EDSS $\geq$4.0: mean difference: –503 steps/day, 95% CI = –638, –368; $p < 0.001$). The ActiGraph and Flex comparisons yielded similar results.

Discussion

Consistent with previous literature,[13,16,40] the ActiGraph tended to under-record steps when fewer steps were taken (slower walking), in the clinic test and also during the home validation (7 d). Greater differences between ActiGraph and Flex (and Flex2) were observed in individuals taking greater steps/day, with lower disability, who were men, and were diagnosed with Relapsing MS. The difference was not related to disability status, although a larger sample may show a small negative differential bias with increasing EDSS scores, independent of daily steps. In light of these results,
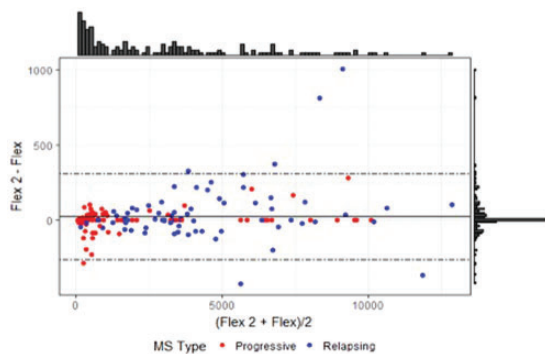

Figure 2. Flex2 and Flex (Bland–Altman Plot). The solid line represents the mean difference between the measures and the dashed horizontal lines represent mean difference ±2 standard deviations. Each solid point corresponds to a separate individual, for a separate day. The solid lighter grey line indicated where '0' difference between the measures would lie. The grey dashed lines are the 95% limits of agreement. The histogram on the top and right depicts the distribution of the data.

researchers should be cautious when choosing what device to use for measuring step count in people with faster walking speeds.

Our previously reported results on Group 1 ($n = 20$) illustrated that Flex recorded 1132 steps/d more than ActiGraph (95% CI = 500–1763).[24] However, the mean difference was small (*c.* 70 steps) for days with low steps (<3000 steps) and increased (*c.* 2400 steps difference) as daily steps increased (>3000 steps/day).[24] The current analysis confirms these results.

This study describes a general method for corroborating measurements of daily step counts captured by remote activity monitoring. The correlation between the Flex and Flex2 step counts demonstrates that step-count data captured by these devices is essentially identical (ICC = 1.0, $p < 0.01$) and can be used interchangeably for remote step count monitoring in individuals with MS. It is likely that other devices using the identical tri-axial accelerometer will yield similar results. However, because manufacturers do not provide hardware or algorithm updates to consumers, it is necessary to establish methods to compare newer to older models thereby ensuring data quality and integrity.

Because prior research indicated greater percentage use of the wrist-worn accelerometers compared to a waist-worn ActiGraph,[22,24] we opted for a 'blocked' times approach to compare the devices and remove potential bias from individuals wearing one device more. Other studies have used a summative approach,[13] however here we selected the blocked approach to favour actual wear-time when all devices were being used concurrently.

The difference between the Flex and Flex2 was negligible (the Flex2 recorded an average increase of 25

Table 3. Interclass correlation coefficients (ICCs) and average differences in steps recorded between measurement techniques during seven days of home activity monitoring (group 2).

	ActiGraph – Flex2	Flex2 – Flex	ActiGraph – Flex
Sample size (*n*)	22	18	20
ICC	0.98	1.00	0.98
(95% confidence interval)	(0.97, 0.99)	(1.00, 1.00)	(0.97, 0.98)
Average difference in steps recorded per device[a]	–808	25	–873

[a]A negative value signifies that the first step counting technique (in the header) measured fewer steps; conversely, a positive sign denotes that the first technique measures more steps than the comparator·

Table 4. Table of Bland–Altman results for all participants and separated into multiple sclerosis subtype (relapsing and progressive) in Group 2.

	Bland–Altman (correlation within Subject)			% within 95% limits of agreement
	Mean difference (95% confidence intervals)	p-value[a]	Difference ~ average (slope, p-value)	(% within modified 95% limits of agreement)[b]
Entire cohort				
ActiGraph v. Flex2	–807.8 (–945.7, –669.9)	<0.01	–0.136, <0.001	94 (96)
ActiGraph v. Flex	–873.4 (–1022.8, –724.1)	<0.01	–0.122, <0.001	93 (96)
Flex2 v. Flex	25.0 (1.7, 48.0)	0.036	0.013, 0.001	95 (92)
Progressive MS				
ActiGraph v. Flex2	–545 (–722, –369)	<0.01	–0.075, 0.03	91 (91)
ActiGraph v. Flex	–625 (–818 –432)	<0.01	–0.06, 0.111	92 (92)
Flex2 v. Flex	3 (–14, 20)	0.72	0.008, <0.01	93 (89)
Relapsing MS				
ActiGraph v. Flex2	–1000 (–1191, –809)	<0.01	–0.151, <0.001	93 (97)
ActiGraph v. Flex	–1066 (–1276, –857)	<0.01	–0.147, <0.01	94 (97)
Flex2 v. Flex	49 (4, 93)	0.034	0.015, 0.071	95 (91)

[a]p-value from paired t-test.
[b]Modified Bland–Altman was used when there was any relationship between the mean and standard deviation of the differences and the magnitude of the measurements.

steps/d than the Flex). However, both Flex and Flex2 wrist-worn devices recorded a significantly greater number of steps than the waist-worn ActiGraph GT3X (on average an increase of 873 and 808 steps/d, respectively). Notably, 96% of individuals measured steps were within the 95% CIs. This observation indicates that wrist-worn and waist-worn devices are measuring the same activity but are judging what constitutes a valid step differently, with the ActiGraph being more conservative. A study comparing the ActiGraph (GT3X+) and the StepWatch observed high accuracy for both devices at fast and comfortable walking speeds over a modified 6-Minute Walk Test, but showed discrepancies (ActiGraph under-recording compared to the StepWatch) at slower speeds and greater ambulatory disability.[13]

The question remains: is Fitbit over-counting or is ActiGraph under-recording? For in-clinic assessments, what constitutes a step is relatively obvious and the participant is fully aware that they are being observed. Walking at home is less predictable in ways that may affect how steps are counted. It is possible that the ActiGraph may not have recorded some activity as full steps whereas the Flex and Flex2 did. The Flex and Flex2, on the other hand, may record other body movements as steps, because these devices are worn around the wrist. Thus the term 'step count' is potentially misleading. It is possible that wrist-worn devices compute higher daily steps compared to the waist-worn ActiGraph because they count movements that may not constitute full-length strides as steps. If so, then it is possible that wrist-worn devices may be more sensitive, and possibly more accurate for overall daily physical activity and function. Weikert et al., proposed that ActiGraph measures 'walking mobility' rather than *physical activity* when worn around the waist in individuals with MS.[41]

These data illustrate the importance of disease and clinical context specific validation of activity monitoring devices for research and clinical care. There is a growing trend of using wearables for health and remote objective monitoring to obtain large amounts of data from people in their daily life. Data from people outside of the clinic-setting give a more holistic view of their function, and may provide clinical trials with more sensitive measures of real-world disability worsening potentially over shorter study durations than traditional performance-based or patient-reported measures. Further study is ongoing to address what other information can be obtained from these wearable devices (i.e. min/min or active v. sedentary bouts) and that may be useful to evaluate activity in the home environment. For example, changes outside of a pre-determined limit could

serve as an early warning sign perhaps prompting the healthcare team to action. To address the discrepancy between the Flex devices and the ActiGraph in greater steps/d and lower disability, studies using other, perhaps more sensitive activity monitors (i.e. StepWatch) are needed.

Study limitations

Although disability score was not found to have a sytematic effect on differential bias, given the sample size larger studies should confirm these results in a similarly wide range of disability. This would increase generalizability across the spectrum. Secondly, there may have been some 'reactivity' due to the short validation time (7 d), although we did not observe any reactivity in our larger cohort study when wrist-worn devices were introduced.[22] Thirdly, the waist-worn ActiGraph appears to be less sensitive to measurement of steps in more disabled persons and perhaps ankle-worn devices (such as StepWatch, which was not used in this study) could provide another perspective on what wrist-worn devices are actually measuring. For the goal of this analysis, the block-time approach was specifically selected in order to compare measurement of all devices when they were actually being worn simultaneously. However, the block-time approach does not account for differences in total daily device usage and therefore would not capture differences in wearability over the course of a valid day that could bias results depending on the research question. As a consequence, comparison of steps using the block time method to the daily average method may not be easily interpretable.

Conclusions

This manuscript presents a cross-validation methodology for remote activity monitoring devices. The Flex and Flex2 models of consumer-grade accelerometer, Fitbit, can be used interchangeably, however, caution should be exercised when using these devices with people who walk many steps per day or have lower levels of disability. Wearable devices, marketed for consumers rather than research, can be useful for longer-term studies as adherence is likely to be higher, albeit at the potential cost of 'exact step count' and detailed gait analysis.

ORCID iD

Valerie J Block ⓘ https://orcid.org/0000-0001-6199-5484

Supplemental Material

Supplemental material for this article is available online.

References

1. Hauser SL and Goodin DS. Multiple Sclerosis and Other Demyelinating Diseases. In: Longo DL, Fauci AS, Kasper DL, et al. (eds). *Harrison's Principles of Internal Medicine*. 18th ed. New York: McGraw-Hill Global Education Holdings, LLC, 2012, p. 33.
2. Compston A and Coles A. Multiple sclerosis. *Lancet* 2008; 372: 1502–1517.
3. Dlugonski D, Pilutti LA, Sandroff BM, et al. Steps per day among persons with multiple sclerosis: variation by demographic, clinical, and device characteristics. *Arch Phys Medicine Rehabil* 2013; 94: 1534–1539.
4. Klaren RE, Motl RW, Dlugonski D, et al. Objectively quantified physical activity in persons with multiple

sclerosis. *Arch Phys Medicine Rehabil* 2013; 94: 2342–2348.

5. Motl RW, Learmonth YC, Pilutti LA, et al. Validity of minimal clinically important difference values for the multiple sclerosis walking scale-12? *Eur Neurol* 2014; 71: 196–202.

6. Motl RW, McAuley E, Wynn D, et al. Lifestyle physical activity and walking impairment over time in relapsing-remitting multiple sclerosis: results from a panel study. *Am J Phys Med Rehabil* 2011; 90: 372–379.

7. Motl RW, Pilutti LA, Learmonth YC, et al. Clinical importance of steps taken per day among persons with multiple sclerosis. *PloS One* 2013; 8: e73247.

8. Lavie CJ and Milani RV. Cardiac rehabilitation and exercise training in secondary coronary heart disease prevention. *Prog Cardiovasc Dis* 2011; 53: 397–403.

9. Sandroff BM, Klaren RE and Motl RW. Relationships among physical inactivity, deconditioning, and walking impairment in persons with multiple sclerosis. *JNPT* 2015; 39: 103–110.

10. Busse ME, Pearson OR, Van Deursen R, et al. Quantified measurement of activity provides insight into motor function and recovery in neurological disease. *J Neurol Neurosurg Psychiatry* 2004; 75: 884–888.

11. Motl R, Sandroff B and Sosnoff J. Commercially available accelerometry as an ecologically valid measure of ambulation in individuals with multiple sclerosis. *Expert Rev Neurother* 2012; 12: 1079–1088.

12. Block VAJ, Pitsch E, Tahir P, et al. remote physical activity monitoring in neurological disease: a systematic review. *PloS One* 2016; 11: e0154335.

13. Sandroff B, Motl R, Pilutti L, et al. Accuracy of StepWatch™ and ActiGraph accelerometers for measuring steps taken among persons with multiple sclerosis. *PLoS One* 2014; 9(4): e93511.

14. Takacs J, Pollock C, Guenther J, et al. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport* 2014; 17: 496–500.

15. Tully MA, McBride C, Heron L, et al. The validation of Fibit Zip physical activity monitor as a measure of free-living physical activity. *BMC Res Notes* 2014; 7: 952.

16. 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: 1476–1485.

17. Diaz KM, Krupka DJ, Chang MJ, et al. Fitbit(R): An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol* 2015; 185: 138–140.

18. Balto JM, Kinnett-Hopkins DL and Motl RW. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. *Mult Scler J Exp Transl Clin* 2016; 2: 2055217316634754.

19. Fitbit. Fitbit Official Site, https://help.fitbit.com/articles/en_US/Help_article/2087/?q=flex+2&l=en_US&fs=Search&pn=1 (2019, accessed 11 November 2019).

20. Marino FE. Heat reactions in multiple sclerosis: an overlooked paradigm in the study of comparative fatigue. *Int J Hyperthermia* 2009; 25: 34–40.

21. Opara JA, Brola W, Wylegala AA, et al. Uhthoff's phenomenon 125 years later: what do we know today? *J Med Life* 2016; 9: 101–105.

22. Block VJ, Lizée A, Crabtree-Hartman E, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *J Neurol* 2016: 1–11.

23. Block VJ, Bove R, Zhao C, et al. Association of continuous assessment of step count by remote monitoring with disability progression among adults with multiple sclerosis. *JAMA Network Open* 2019; 2: e190570–e190570.

24. Block VJ. *FITBIT remote-monitoring in multiple sclerosis: FITriMS study.* PhD Thesis, University of California San Francisco, USA.

25. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011; 69: 292–302.

26. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014; 83: 278–286.

27. Kurtzke J. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.

28. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med* 2013; 34: 975–982.

29. Motl RW, Snook EM, Agiovlasitis S, et al. Calibration of accelerometer output for ambulatory adults with multiple sclerosis. *Arch Phys Med Rehabil* 2009; 90: 1778–1784.

30. ActiGraph. User Guide ActiGraph wGT3X-BT and ActiLife, Actigraph, 2016.

31. Choi L, Liu Z, Matthews CE, et al. Validation of accelerometer wear and nonwear time classification algorithm. *Med Science Sports Exerc* 2011; 43: 357–364.

32. Troiano (2007) Wear Time Validation Parameters, https://actigraphcorp.force.com/support/s/article/Troiano-2007-Wear-Time-Validation-Parameters (2019, accessed 11 November 2019).

33. Arguello D, Andersen K, Morton A, et al. Validity of proximity sensor-based wear-time detection using the ActiGraph GT9X. *J Sports Sci* 2018; 36: 1502–1507.

34. Motl RW, Zhu W, Park Y, et al. Reliability of scores from physical activity monitors in adults with multiple sclerosis. *APAQ* 2007; 24: 245–253.

35. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.

36. Bland JM and Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007; 17: 571–582.

37. Myles PS and Cui J. Using the Bland–Altman method to measure agreement with repeated measures. *BJA* 2007; 99: 309–311.

38. Leave-One-Out Cross-Validation. In: Sammut C and Webb GI (eds) *Encyclopedia of Machine Learning*. Boston, MA: Springer, 2010, pp.600–601.

39. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org (2019, accessed 11 November 2019).

40. Singh AK, Farmer C, Van Den Berg ML, et al. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. *Disabil Health J* 2016; 9: 320–323.

41. Weikert M, Suh Y, Lane A, et al. Accelerometry is associated with walking mobility, not physical activity, in persons with multiple sclerosis. *Med Eng Phys* 2012; 34: 590–597.