Title: Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica

Authors: Anish Chandrasekar, Elizabeth M.A. Hensor, Sarah L. Mackie, Michael R. Backhouse, Emma Harris

PII: S0966-6362(18)30046-8
DOI: https://doi.org/10.1016/j.gaitpost.2018.01.035
Reference: GAIPOS 5949

To appear in: *Gait & Posture*

Received date: 6-7-2017
Revised date: 23-1-2018
Accepted date: 28-1-2018

Please cite this article as: Chandrasekar Anish, Hensor Elizabeth MA, Mackie Sarah L, Backhouse Michael R, Harris Emma.Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica.*Gait and Posture* https://doi.org/10.1016/j.gaitpost.2018.01.035

Title: Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica

Authors: Anish Chandrasekar[1], Elizabeth M.A. Hensor[2,3], Sarah L. Mackie[2,3], Michael R. Backhouse[2,3], Emma Harris[2,4]

Affiliations:

[1] School of Medicine, University of Leeds, Leeds, LS2 9JT, United Kingdom (email: um13a4c@leeds.ac.uk)

[2] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, LS7 4SA, United Kingdom (email: M.R.Backhouse@leeds.ac.uk; E.M.A.Hensor@leeds.ac.uk; s.l.mackie@leeds.ac.uk)

[3] NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, LS7 4SA, United Kingdom

[4] School of Human and Health Sciences, University of Huddersfield, Huddersfield, HD1 3DH, United Kingdom (E.Harris@hud.ac.uk)

Correspondence to: Michael Backhouse Clinical Biomechanics and Physical Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, 2nd Floor Chapel Allerton Hospital, Harehills Lane, Leeds, LS7 4SA, United Kingdom.

Email: M.R.Backhouse@leeds.ac.uk

Tel: +44(0) 113 3924926

Abstract word count: **262**

Current word count: **2999** (excluding reference list, abstract, tables and figures)

Research Highlights:

- Results suggest the activity monitors are less accurate in PMR than healthy people
- Fitbit-Zip appears to fairly accurately measure step-count during walking in PMR
- Results suggest Actigraph with LFE was the most accurate device for stair ascent
- Device accuracy decreased in PMR patients with higher functional impairment
- Actigraph's LFE filter increased device accuracy in controlled conditions

Abstract

Background

Activity monitors provide objective measurements of physical activity, however, the accuracy of these devices in people with polymyalgia rheumatica (PMR) is unknown. Therefore, this study aimed to obtain preliminary evidence of the accuracy of two activity monitors and explore if clinical and gait-related factors altered device accuracy in people with PMR.

Methods

The *ActiGraph* with low frequency extension (+LFE) and standard (-LFE) algorithms, *Fitbit-Zip* (waist) and *Fitbit-Zip* (shirt) were concurrently tested using a two-minute walk test (2MWT) and stairs test in 27 people with PMR currently treated with prednisolone. To determine accuracy, activity monitor step-count was compared to a gold-standard step-count (GSSC; calculated from video recording) using Bland-Altman plots.

Results

The *Fitbit-Zip* (waist) achieved closest agreement to the GSSC for the 2MWT (mean bias (95%CI): 10 (-3, 23); 95%LOA: -55, 74). The *ActiGraph* (+LFE) achieved closest agreement to the GSSC for the stairs test (mean bias (95%CI): 0 (-1, 1); 95%LOA: -5, 5). The *ActiGraph* (-LFE) performed poorly in both tests. All devices demonstrated reduced accuracy in participants with lower gait velocity, reduced stride length, longer double-limb support phase and greater self-reported functional impairment.

Conclusion

Our preliminary results suggest that in controlled conditions, the *Fitbit-Zip* fairly accurately measures step-count during walking in people with PMR receiving treatment. However, device error was greater than data published in healthy people. The *ActiGraph* may not be recommended without activation of the LFE. We identified clinical and gait-related factors associated with higher levels of functional impairment that reduced device accuracy. Further work is required to evaluate the validity of the activity monitors in field conditions.

Key words: Activity monitoring, Physical activity, Step-count, Polymyalgia rheumatica, Accelerometers, Device comparison

1. Introduction

Promoting physical activity is a key aspect of public health policy [1], and advice on increasing physical activity is commonly given during clinical consultations. Despite this, an estimated one third of adults globally do not meet recommended exercise guidelines; inactivity is even higher in older adults and people with chronic disease [1]. Polymyalgia rheumatica (PMR) is a common chronic, inflammatory musculoskeletal disease that usually affects adults over the age of 50 [2, 3]. PMR is treated with systemic glucocorticoid therapy, which often causes proximal myopathy and weight gain, exacerbating the effect of pre-existing osteoarthritis on gait, posture and mobility [4]. Physical activity and exercise are therefore recommended to mitigate physical deconditioning and promote bone health [5].

Wearable accelerometers that measure step-count are popular devices for self-tracking physical activity and are a useful self-management support tool for increasing physical activity in daily life [6]. The ability of the user to create goals and track step-counts in real time is thought to positively change exercise behaviour [7], which might be useful for people with PMR. Popular consumer activity monitors such as the *Fitbit-Zip* and research

accelerometers such as the *ActiGraph*, can accurately measure step-count in healthy young [8-11] and older [12] adults. A Low Frequency Extension (LFE) filter was developed for the *ActiGraph* to increase sensitivity and improve step-count accuracy during lower-intensity movements. However, the algorithms used to calculate step-count were developed on young, healthy people. Many clinical conditions are associated with abnormalities of gait and posture that have been shown to affect the accuracy of these algorithms [12-17], yet step-count accuracy has not been tested in people with PMR. Prior to using activity monitors to measure the physical activity of this population and developing subsequent interventions to increase physical activity, it is important to explore the validity of activity monitors in this target population.

Movement type and direction also affect the accuracy of activity monitors to record steps [18]. However, step-count accuracy has most commonly been assessed during rhythmic movement with forward motion, such as walking or running. Walking up stairs is a common activity which involves different movement patterns to walking on a flat surface but has rarely been explored in studies measuring step-count accuracy [8]. Furthermore, the effect of *ActiGraph's* LFE filter on step-count accuracy has not been tested during stair climbing.

The purpose of this study was to obtain preliminary evidence of the accuracy of two commonly used activity monitors in patients with PMR who are treated with systemic glucocorticoid therapy. The performance of the *Fitbit-Zip* and *ActiGraph* devices were compared to direct measurement using video-observation, in a group of patients on glucocorticoid treatment for PMR. Device accuracy was assessed during walking on flat ground and ascending stairs, as examples of everyday activities with different movement patterns. The effect of *Fitbit-Zip* monitor location *(chest vs. waist)* and the application of the *ActiGraph* LFE filter on step-count accuracy was also explored. Lastly, in exploratory

analyses we examined associations between gait-related and clinical variables and device error.

2. Methods

2.1. Participants

Patients were recruited from rheumatology outpatient clinics between 2015 and 2016. All patients had a consultant-confirmed diagnosis of PMR according to clinical practice guidelines [19], were treated with systemic glucocorticoids (prednisolone) for at least one month, and agreed to walk for at least two minutes. The Leeds NHS Research Ethics Committee granted ethical approval (Ref:14/YH/1147) and all patients provided written consent prior to participation in accordance with the Declaration of Helsinki.

2.2. Activity Monitors

The *Fitbit-Zip* (Fitbit, Inc, San Francisco, USA) is a small, commercially-available physical activity monitor worn at the waist or midline of the shirt. Step-counts are displayed on the LCD screen. The *ActiGraph-GT3X+* (ActiGraph LLC, Pensacola, USA) is a slightly larger research device most commonly worn at the hip. Raw data are downloaded and processed using *ActiLife.6* software on a computer, generating output parameters including step-count. Both devices use the integrated 3-axis accelerometer, which quantifies acceleration, to calculate step-counts using manufacturers' algorithms. Individual gait parameters such as stride length are not considered in these algorithms.

2.3. Study Protocol

On the assessment day, participants' height and weight were measured and participants completed patient-reported outcome questionnaires; the Stanford Health Assessment Questionnaire Disability Index (HAQ-DI) [20] and the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) questionnaire [21]. Gait velocity, stride length and duration of double-limb support phase (DLSP) were measured while the

participants walked at a self-selected normal pace, using the previously validated GAITRite walkway [22]. Participants' leg length was measured and entered into the GAITRite software (CIR Systems Inc., Sparta, USA) to enable the calculation of the gait parameters, as per the manufacturer's instructions. Serum C-reactive protein (CRP) was measured within one week of the study visit, as part of the clinical assessment for PMR disease activity.

Participants completed a two-minute-walk-test (2MWT) and stairs test whilst simultaneously wearing three devices (one *ActiGraph* and two *Fitbit-Zip* monitors). The *ActiGraph* and first *Fitbit-Zip* were mounted side-by-side on participants' right hip using an elastic waistband, whilst the second *Fitbit-Zip* was attached to the midline of the shirt. Throughout the tests, participants were asked to walk at their normal pace and use any assistive devices required. Participants were allowed to rest or stop the test where necessary. To avoid recording extra steps, participants were asked to remain still at the end of each test whilst step-count data were transcribed from the devices.

The 2MWT evaluated step-count accuracy of devices whilst freely walking on flat ground. Participants walked around the gait laboratory, using a self-dictated route for a total of two minutes.

The stairs test evaluated step-count accuracy of devices during stair ascent. Participants were seated at the bottom of a 12-step staircase in the hospital, and asked to ascend the stairs and sit down on a chair placed at the top of the staircase. Throughout both tests, video footage was recorded to accurately capture steps and enable later calculation of the gold standard step-count (GSSC). Different *ActiGraph* devices were used for each test for practicality reasons, as tests were performed sequentially and the staircase was not adjacent to the gait laboratory.

2.4 Data processing and analysis

In the absence of existing data on which to base a power calculation the sample size was based on published rules of thumb for preliminary studies, which recommend up to 30 patients be recruited [23]. This is comparable to previous validation studies that have used sample sizes between 21 to 32 [9, 12, 15, 24]. *Fitbit-Zip* step-count was calculated as the difference between steps displayed on the device at the start and end of each test. The *ActiGraph* devices recorded step-counts at 30Hz for two minutes during the 2MWT and one minute for the stairs test. Step-count data were analysed in 60-second epochs, with and without the LFE filter, using ActiLife software (v.6.13.2, ActiGraph LLC). The LFE aims to improve step-count accuracy by changing the band-pass filter settings to enable low-intensity activities to be recorded. Preliminary evidence of step-count accuracy was obtained for four device conditions: *Fitbit-Zip* (waist), *Fitbit-Zip* (shirt), *ActiGraph* (-LFE) and *ActiGraph* (+LFE).

To calculate the GSSC of each test, two researchers (A.C and E.H) replayed videos of each participant at half-speed, and individually counted the number of steps three times whilst blinded to the other's count, to yield a mean value. A step was defined as a 'transition from the first point of contact of one foot with the ground, through to the first point of contact with a contralateral foot'. As such, participants' feet did not need to completely leave the ground to be counted as a step, in order to account for shuffling. Inter- and intra-rater reliability (IRR) of visual count from video-recordings were assessed concurrently using fixed-effects intra-class correlation coefficient (ICC), calculated according to Eliasziw *et al* [25]. An ICC<0.20 was considered *slight*, 0.21-0.40 *fair*, 0.41-0.60 *moderate*, 0.61-0.80 *substantial* and 0.81-1.00 *almost perfect* [25].

Bland-Altman plots measured agreement between activity monitors and the GSSC [26, 27]. Plots described accuracy of each device by the mean step-count difference (bias) and 95% limits of agreement (95%LOA). The more narrow the 95%LOA and closer the bias to

the line of equality (Y=0), the more accurate the device, and vice versa. As recommended by Bland *et al* [27], outliers were not excluded during analysis. If the difference between measurements was found to be related to the mean measurement, the regression method was used to calculate the 95%LOA. If the absolute values of the residuals from the regression model were found to be related to the mean measurement at $p<.05$, the 95%LOA were adjusted accordingly [27]. Mean absolute percentage error (MAPE) was also calculated.

For consistency non-parametric Spearman's rank correlation was used to assess all associations between step-count bias and demographic, clinical and gait parameters because many of the interval-scaled outcomes were highly skewed (Body Mass Index (BMI), steroid dose, CRP, stride length, gait velocity) and the patient-reported outcomes (HAQ-DI, FACIT-Fatigue) were ordinal-scaled. Absolute rho values ≥0.3 were considered substantive. All analyses were performed using Stata software (Stata v.14.1, StataCorp, College Station, TX, USA).

4. Results

4.1 Participant Characteristics and associations between gait-related and clinical variables

Demographic and clinical characteristics of the 27 patients who took part in the study are presented in Table 1. Three patients also had giant cell arteritis (GCA) and two had a diagnosis of osteoarthritis (OA). One patient used two walking canes during both tests and nine used the handrail during stair-ascent. Age, HAQ-DI scores and CRP were found to have substantive and statistically significant (rho>0.4, $p < 0.05$) associations with DLSP, stride length and gait velocity. In addition, DLSP demonstrated association to the same degree with BMI. Several other associations (0.40>rho>0.30) were also found (Table 2).

4.2 Step-count accuracy

All participants except one completed both tests; due to fatigue this participant was only able to complete 30s of the 2MWT and was unable to complete the stairs test. This participant's 2MWT data was included in analysis as recommended by Bland *et al.,* [27]. Step-count data is detailed in the supplementary file (supplementary Table 1). For the gold-standard video assessment, IRR between observers was almost perfect for both tests (ICC>0.99). All monitors systematically underestimated step-counts during the 2MWT and stairs test, as demonstrated by the mean bias falling above the line of equality (Y=0; Figures 1 and 2). During the 2MWT, the *Fitbit-Zip* (waist) achieved the closest agreement to the GSSC, followed by the *Fitbit-Zip* (shirt), *ActiGraph* (+LFE) and *ActiGraph* (-LFE) (Figure 1, Table 3). Conversely, during the stairs test, the *ActiGraph* (+LFE) achieved the closest agreement to the GSSC, followed by the *Fitbit-Zip* (waist) and *ActiGraph* (-LFE), whilst the *Fitbit-Zip* (shirt) attained the poorest agreement (Figure 2, Table 3).

Although the *Fitbit-zip* (shirt) performed well during the 2MWT, it was less accurate during the stairs test. This was largely due to two anomalous results where the *Fitbit-zip* (shirt) recorded 207 steps and zero steps on two different occasions. It is unclear whether this was device or human error. Sensitivity analysis excluding these two results improved the step-count accuracy of the *Fitbit-zip* (shirt) from 64% mean error to 10% mean error (supplementary Table 2). Apart from these two anomalous results, the position of the Fitbit had little effect on step-count accuracy.

ActiGraph's LFE algorithm had a large effect on device accuracy; the default setting (-LFE) reduced device accuracy during both tests and increased step-count underestimation when fewer steps were taken (higher mean bias and wider 95%LOA, Table 3). When LFE was activated, step-count agreement was improved and more comparable to the *Fitbit-zip* (Table 3), although mean step-count bias was still larger than the *Fitbit-zip* for the 2MWT.

Bland-Altman plots (Figure 1) demonstrated that participants who took fewer steps during the 2MWT recorded greater step-count underestimation in all devices, especially the *ActiGraph* (-LFE). For all devices, step-count bias correlated negatively with gait velocity (rho=-.37 to -.58) and stride length (rho=-.31 to -.72) and positively with HAQ-DI scores (rho=.32 to .57) and duration of DLSP (rho=.33 to .61; Table 2).

5. Discussion

This study was the first to measure the accuracy of accelerometer-based activity monitors in people with PMR. Based on close agreement with the video-recorded gold-standard, our preliminary results suggest that the *Fitbit-Zip* (waist) was the most accurate measure of step-count whilst walking on flat ground in people with PMR, closely followed by the *Fitbit-Zip* (shirt). In contrast, the *ActiGraph* (+LFE) was the most accurate device during stair ascent, followed by the *Fitbit-Zip* (waist) then *ActiGraph* (-LFE). The *Fitbit-Zip* (shirt) was the poorest performing when all participants were included. However, when two anomalous results were excluded, its accuracy was similar to the *Fitbit-Zip* (waist) and *ActiGraph* (+LFE). Altered lower limb kinematics may explain the differences in step-count accuracy between level walking and stair climbing; larger knee and hip flexion and no heel contact with the ground are observed during stair ascent. This may have resulted in larger acceleration, which devices are more able to detect. [28]. The *ActiGraph* (-LFE) was more accurate than *Fitbit-Zip* (shirt) on the stairs test when all patients were included; however, when two anomalous results were excluded, it had the greatest mean bias and widest limits of agreement of all tests.

Consistent with previous findings in older adults [12], our results also suggest the Fitbit-Zip to be more accurate than ActiGraph for recording step-count during level walking. Although the *FitBit-zip* (waist) performed better than the other device conditions, it still performed relatively poorly in PMR patients (Table 3), compared to published data from

healthy young adults [8]. Similarly, while the *ActiGraph* (+LFE) performed best during stair ascent, the step-count mean error was considered large at 8%.

Step-count accuracy was lower in participants who took fewer steps during the 2MWT, walked slower, had shorter stride length, longer double-limb support phase and greater self-reported functional impairment. Previous studies have reported similar findings in older adults [12, 14] although in one young, healthy population, walking speed did not affect step-count accuracy [8]. This was likely because the population studied did not have gait abnormalities and the lowest speed tested was 90 cm/s [8]. Indeed, *FitBit-zip* (waist) step-count accuracy is reported to be as high as 98% at walking speeds of 90 cm/s but as low as 47% at 50 cm/s [13].

Step-count error at slower walking speeds in PMR might be explained by device algorithms that convert raw acceleration profiles into step-counts, or inaccuracies in device hardware itself [29]. Accelerations during movement must meet/surpass a threshold in order for a step to be recognised. However, the algorithms were based on kinetic gait characteristics of healthy adults [30]. At slower walking speeds, the acceleration amplitude is lower [31], and may be insufficient to surpass the acceleration threshold required to register a valid step [32]. This mechanism may explain why step-count error was exacerbated in ActiGraph devices without activation of LFE during level walking and stair climbing.

ActiGraph's LFE filter reduces the acceleration threshold by expanding the bandwidth of the normal software filtering algorithm [33]. This increases the lower limit of the acceleration cut-off, which increases ActiGraph's sensitivity to low-intensity movements that was observed in participants this study; (the median gait velocity was slower than published data on age-matched healthy people [34]). During both tests in this study, the Actigraph's step-count accuracy was improved by activating LFE. Although this is consistent with previous studies conducted in controlled conditions [33, 35], the activation of LFE in free-

living conditions in young and older adults overestimates step-count [33, 36]. Therefore, further research is required to evaluate the validity of these activity monitors in PMR in free-living conditions.

5.1 Strengths and Limitations:

This study tested two commonly used activity monitors in a well-defined clinical cohort of PMR patients, using a protocol previously described in similar studies [12, 15]. Error attributable to intra-subject variation was minimised because the devices were tested concurrently. The controlled laboratory environment of our study facilitated a fair, repeatable objective evaluation of activity monitors. Although the study evaluated the activity monitors in both walking along a flat surface and on stairs, the controlled conditions may limit external validity. Therefore, it is important that the activity monitors are further tested in free-living conditions outside of the laboratory environment.

Bland-Altman plots enabled separation of systematic error from the activity monitor and random error from unknown/unpredictable changes in the experiment [37], such as human error. It is therefore an appropriate method for testing agreement and thus, interchangeability between two methods of measuring step-count (i.e. GSSC vs. activity monitor) [27]. Validation studies such as this may benefit from implementing *a priori* criteria that states the minimal clinically important difference in step-count between accelerometer and GSSC, based on the clinical and biological traits of people with PMR. In theory this would further optimise the Bland-Altman system [26] and thus statistical validity. Furthermore, we have explored clinical factors influencing accuracy that might introduce systematic bias into analyses of clinical studies in which physical activity is incorporated as either a covariate or an outcome. Due to the small sample size in this study, our results are preliminary and should be confirmed in a further study with a larger sample size.

Females are more likely to develop PMR than males [3], which is reflective of the gender distribution in this study. Gender based differences in gait patterns have been previously reported [38, 39]. Although this could potentially affect the accuracy of activity monitors, exploring this possibility was beyond the scope of this study.

6.0 Conclusion:

Our preliminary results suggest that in controlled conditions, the *Fitbit-Zip* is a fairly accurate measure of step-count during walking in people with PMR receiving glucocorticoid treatment. However, when compared to validation data published in healthy people, the devices were not as accurate for measuring steps. Further research is needed to determine its accuracy during stair ascent when worn on the chest. Step-count accuracy of *ActiGraph* devices appears to be improved with activation of the LFE filter. Increasing functional impairment or gait abnormalities reduced step-count accuracy in all devices. Further work is required to evaluate the validity of these activity monitors in field conditions.

6.0 Conflicts of interest

The authors have no conflicts of interest to report.

Acknowledgements: This research was supported by a European Research Council grant (PRECORT- 250271) and the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would like to thank Dr Louise Sorensen for her assistance with data collection.

7.0 References:

[1] World Health Organisation. Physical activity strategy for the WHO European Region 2016–2025. In: Europe WHOROf, editor. Vilnius, Lithuania2015.
[2] Mackie SL, Mallen CD. Polymyalgia rheumatica. BMJ : British Medical Journal. 2013;347.
[3] Smeeth L, Cook C, Hall AJ. Incidence of diagnosed polymyalgia rheumatica and temporal arteritis in the United Kingdom, 1990-2001. Ann Rheum Dis. 2006;65:1093-8.
[4] Hutchings A, Hollywood J, Lamping DL, Pease CT, Chakravarty K, Silverman B, et al. Clinical outcomes, quality of life, and diagnostic uncertainty in the first year of polymyalgia rheumatica. Arthritis and rheumatism. 2007;57:803-9.
[5] Dejaco C, Singh YP, Perel P, Hutchings A, Camellino D, Mackie S, et al. Current evidence for therapeutic interventions and prognostic factors in polymyalgia rheumatica: a systematic literature review informing the 2015 European League Against Rheumatism/American College of Rheumatology recommendations for the management of polymyalgia rheumatica. Ann Rheum Dis. 2015;74:1808-17.
[6] de Bruin ED, Hartmann A, Uebelhart D, Murer K, Zijlstra W. Wearable systems for monitoring mobility-related activities in older people: a systematic review. Clinical rehabilitation. 2008;22:878-95.
[7] Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC medicine. 2015;13:77.
[8] Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait & Posture. 2016;48:36-41.
[9] Adam Noah J, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol. 2013;37:456-62.
[10] Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit(R): An accurate and reliable device for wireless physical activity tracking. International journal of cardiology. 2015;185:138-40.
[11] Abel MG, Hannon JC, Sell K, Lillie T, Conlin G, Anderson D. Validation of the Kenz Lifecorder EX and ActiGraph GT1M accelerometers for walking and running in adults. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2008;33:1155-64.
[12] Paul SS, Tiedemann A, Hassett LM, Ramsay E, Kirkham C, Chagpar S, et al. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport & Exercise Medicine. 2015;1.
[13] Singh AK, Farmer C, Van Den Berg MLE, Killington M, Barr CJ. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. Disability and Health Journal. 2016;9:320-3.
[14] Storti KL, Pettee KK, Brach JS, Talkowski JB, Richardson CR, Kriska AM. Gait speed and step-count monitor accuracy in community-dwelling older adults. Med Sci Sports Exerc. 2008;40:59-64.
[15] Backhouse MR, Hensor EMA, White D, Keenan A-M, Helliwell PS, Redmond AC. Concurrent validation of activity monitors in patients with rheumatoid arthritis. Clinical Biomechanics (Bristol, Avon). 2013;28:473-9.
[16] Fulk GD, Combs SA, Danks KA, Nirider CD, Raja B, Reisman DS. Accuracy of 2 Activity Monitors in Detecting Steps in People With Stroke and Traumatic Brain Injury. Physical therapy. 2014;94:222-9.

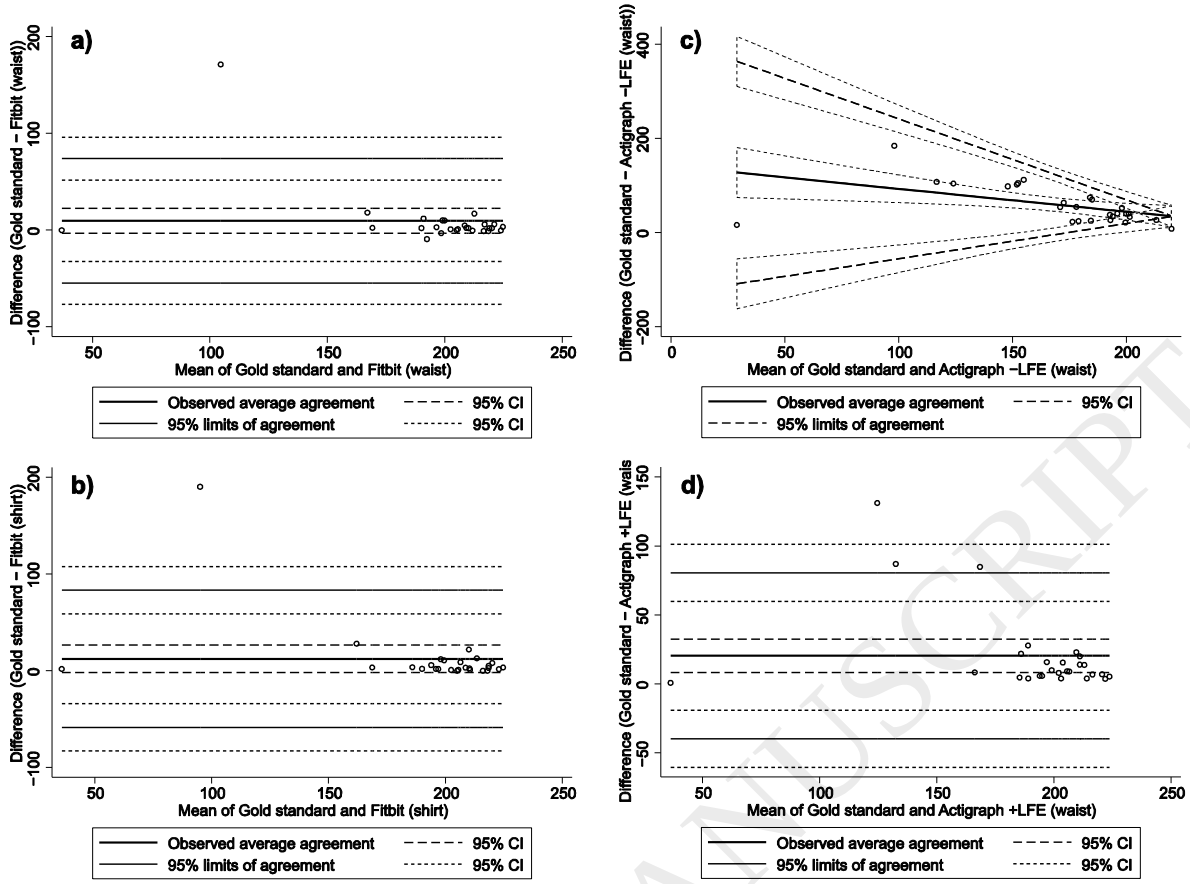
[17] Dijkstra B, Zijlstra W, Scherder E, Kamsma Y. Detection of walking periods and number of steps in older adults and patients with Parkinson's disease: accuracy of a pedometer and an accelerometry-based method. Age and Ageing. 2008;37:436-41.
[18] Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of Activity Monitor Step Detection Is Related to Movement Patterns. Journal of physical activity & health. 2016;13:145-53.
[19] Dasgupta B, Borg FA, Hassan N, Barraclough K, Bourke B, Fulcher J, et al. BSR and BHPR guidelines for the management of polymyalgia rheumatica. Rheumatology (Oxford, England). 2010;49:186-90.
[20] Kalke S, Mukerjee D, Dasgupta B. A study of the health assessment questionnaire to evaluate functional status in polymyalgia rheumatica. Rheumatology (Oxford, England). 2000;39:883-5.
[21] Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. The Journal of rheumatology. 2005;32:811-9.
[22] Menz HB, Latt MD, Tiedemann A, Mun San Kwan M, Lord SR. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. Gait Posture. 2004;20:20-5.
[23] Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. Journal of evaluation in clinical practice. 2004;10:307-12.
[24] Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. International Journal of Behavioral Nutrition and Physical Activity. 2015;12:42.
[25] Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Physical therapy. 1994;74:777-88.
[26] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet (London, England). 1986;1:307-10.
[27] Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical methods in medical research. 1999;8:135-60.
[28] Riener R, Rabuffetti M, Frigo C. Stair ascent and descent at different inclinations. Gait & Posture. 2002;15:32-44.
[29] Feng Y, Wong CK, Janeja V, Kuber R, Mentis HM. Comparison of tri-axial accelerometers step-count accuracy in slow walking conditions. Gait & Posture. 2017;53:11-6.
[30] Tudor-Locke C, Ainsworth BE, Thompson RW, Matthews CE. Comparison of pedometer and accelerometer measures of free-living physical activity. Med Sci Sports Exerc. 2002;34:2045-51.
[31] Lelas JL, Merriman GJ, Riley PO, Kerrigan DC. Predicting peak kinematic and kinetic parameters from gait speed. Gait Posture. 2003;17:106-12.
[32] Taraldsen K, Askim T, Sletvold O, Einarsen EK, Bjastad KG, Indredavik B, et al. Evaluation of a body-worn sensor system to measure physical activity in older people with impaired function. Physical therapy. 2011;91:277-85.
[33] Feito Y, Garner HR, Bassett DR. Evaluation of ActiGraph's low-frequency filter in laboratory and free-living environments. Med Sci Sports Exerc. 2015;47:211-7.
[34] Bohannon RW. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. Age and Ageing. 1997;26:15-9.

[35] Webber SC, St John PD. Comparison of ActiGraph GT3X+ and StepWatch Step Count Accuracy in Geriatric Rehabilitation Patients. Journal of aging and physical activity. 2016;24:451-8.
[36] Wallen MB, Nero H, Franzen E, Hagstromer M. Comparison of two accelerometer filter settings in individuals with Parkinson's disease. Physiological measurement. 2014;35:2287-96.
[37] Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Computers in biology and medicine. 1990;20:337-40.
[38] Chumanov ES, Wall-Scheffler C, Heiderscheit BC. Gender differences in walking and running on level and inclined surfaces. Clin Biomech (Bristol, Avon). 2008;23:1260-8.
[39] Ko SU, Tolea MI, Hausdorff JM, Ferrucci L. Sex-specific differences in gait patterns of healthy older adults: results from the Baltimore Longitudinal Study of Aging. Journal of biomechanics. 2011;44:1974-9.

Figure Captions

Figure 1. Bland-Altman plots measuring and comparing step-count agreement between activity monitors and the gold-standard during the Two-Minute-Walk-Test (2MWT)**.** The more narrow the 95% limits of agreement and closer the bias is to the line of equality, Y=0 (i.e. perfect average agreement) the closer the device agrees to the gold standard step-count (i.e. the more accurate) and vice versa. *Fitbit-Zip* (waist) achieved closest agreement (Figure 1a). All participants completed the 2MWT (n=27).

Figure 2. Bland-Altman plots measuring and comparing step-count agreement between activity monitors and the gold-standard during the Stairs Test. The more narrow the 95% limits of agreement and closer the bias is to the line of equality, Y=0 (i.e. perfect average agreement) the closer the device agrees to the gold-standard step-count (i.e. the more accurate) and vice versa. *Actigraph* (+LFE) achieved closest agreement (Figure 2d), followed closely by *Fitbit-Zip* (waist) (Figure 2a). Twenty six participants completed the stairs test.



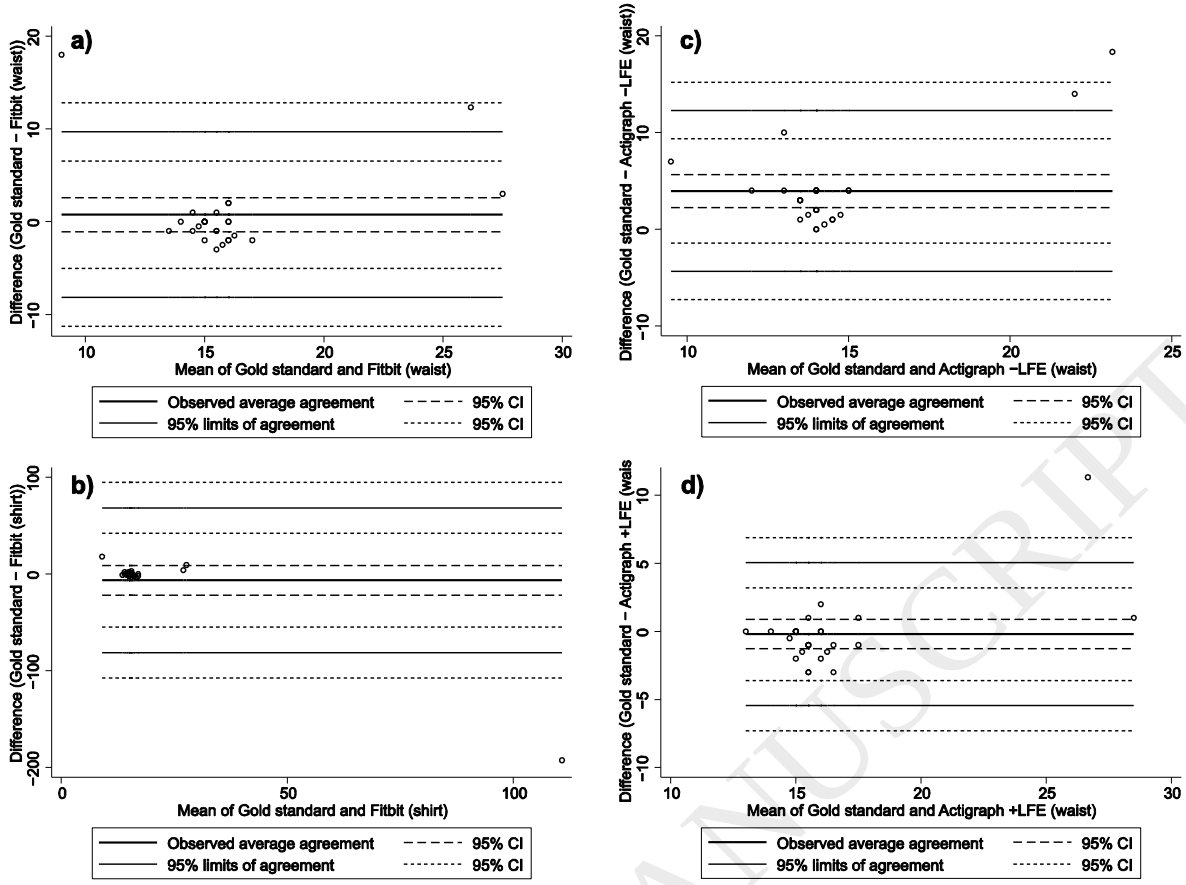


Table Captions

Table 1.

Summary statistics for patient demographics and clinical details.

Key: %GC = percentage of gait cycle, SD = Standard deviation, IQR = Inter-quartile range, BMI = Body Mass Index, CRP = C-Reactive Protein, HAQ-DI = Stanford Health Assessment Questionnaire Disability Index, FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue. The maximum possible HAQ-DI score is 3; higher HAQ-DI scores indicate higher functional impairment. The maximum possible FACIT-Fatigue score is 52; higher FACIT-Fatigue scores indicate lower levels of fatigue. Double support refers to the proportion of the gait cycle spent in double-limb support (DLS), measured as a percentage of the gait cycle. For a normal gait cycle, DLS is approximately 20%. People with mobility issues usually have higher values for DLS.

	Median (IQR)	Mean (SD)	Range	Number of participants (%)
Female	-	-	-	24 (89)
Age (years)	-	69.2 (8.8)	50 - 85	27 (100)
BMI (kg/m^2)	-	28.3 (5.6)	19.9 – 40.4	27 (100)
Gait Velocity (cm/s)	119 (95, 131)	-	46 – 142	27 (100)
Stride length (cm)	125 (111, 140)	-	61 – 163	27 (100)
Double Support (%GC)	25.9 (23.9, 29.8)	-	20.4 – 38.0	27 (100)
Daily steroid dose (mg)	9.0 (5.0, 12.5)	-	1 – 40	27 (100)
CRP (mg/L)	<5 (<5, 17)	-	<5 – 60	25 (92.6)
HAQ-DI	0.63 (0.13, 1.25)	-	0.00 – 2.13	26 (96.3)
FACIT-Fatigue	38 (33, 45)	-	5 – 51	25 (92.6)

Key: %GC = percentage of gait cycle, SD = Standard deviation, IQR = Inter-quartile range, BMI = Body Mass Index, CRP = C-Reactive Protein, HAQ-DI = Stanford Health Assessment Questionnaire Disability Index, FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue. The maximum possible HAQ-DI score is 3; higher HAQ-DI scores indicate higher functional impairment. The maximum possible FACIT-Fatigue score is 52; higher FACIT-Fatigue scores indicate lower levels of fatigue. Double support refers to the proportion of the gait cycle spent in double-limb support (DLS), measured as a percentage of the gait cycle. For a normal gait cycle, DLS is approximately 20%. People with mobility issues usually have higher values for DLS.

Table 2. Spearman's rho associations of factors influencing gait and step-count bias

The first section of this table shows the associations between gait parameters and demographic and clinical parameters whilst the second section shows associations between step-count bias and demographic, clinical and gait parameters. |Rho| ≥0.3 is considered to be substantive association (for reference, threshold rho values for statistical significance in this sample size are ≥ 0.4 $p < 0.05$; ≥-0.33 $p < 0.1$). Values highlighted in bold indicate correlations where substantive association was found.

Key: BMI = Body Mass Index, CRP = C-Reactive Protein, HAQ-DI = Stanford Health Assessment Questionnaire Disability Index, FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue.

	Gait parameters			Two minute walk test (n = 27) step-count bias				Stairs test (n = 26) step-count bias			
	Stride length	Gait velocity	Duration double support phase	*Fitbit-zip*		*Actigraph*		*Fitbit-zip*		*Actigraph*	
				Waist	Shirt	-LFE	+LFE	Waist	Shirt	-LFE	+LFE
Age	**-0.47**	**-0.45**	**0.44**	**.46**	**.58**	**.34**	**.37**	-.04	.10	-.16	-.09
BMI	**-0.37**	**-0.30**	**0.56**	.02	-.08	.26	.01	.04	.02	.20	.03
Steroid dose	-0.23	-0.22	**0.35**	.24	.20	.23	**.43**	.06	.24	-.07	-.16
CRP	**-0.48**	**-0.45**	**0.45**	.25	**.33**	.22	.07	.15	.12	.12	.16
HAQ-DI	**-0.65**	**-0.57**	**0.62**	**.55**	**.46**	**.57**	**.35**	**.32**	**.40**	**.50**	**.41**
FACIT-Fatigue	**0.37**	**0.30**	-0.25	**-.33**	-.08	**-.55**	-.23	-.28	-.23	**-.38**	**-.43**
Stride length				**-.55**	**-.47**	**-.72**	**-.45**	**-.50**	**-.45**	**-.34**	**-.31**
Gait velocity				**-.43**	**-.44**	**-.56**	**-.37**	**-.58**	**-.46**	**-.42**	**-.40**
Duration of double support				**.49**	**.44**	**.61**	**.43**	**.40**	**.33**	**.39**	**.37**

The first section of this table shows the associations between gait parameters and demographic and clinical parameters whilst the second section shows associations between step-count bias and demographic, clinical and gait parameters. |Rho| ≥0.3 is considered to be substantive association (for reference, threshold rho values for statistical significance in this sample size are ≥ 0.4 $p < 0.05$; ≥-0.33 $p < 0.1$). Values highlighted in bold indicate correlations where substantive association was found.

Key: BMI = Body Mass Index, CRP = C-Reactive Protein, HAQ-DI = Stanford Health Assessment Questionnaire Disability Index, FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue.

Table 3. Bland-Altman 95% limits of agreement, bias (mean difference) and mean absolute percentage error for Two Minute Walk Test and Stair test**.**

The table illustrates the key statistics taken from Bland-Altman analysis. Bias indicates the mean difference between the gold standard step-count and device, whilst the 95%LOA demonstrates the upper and lower limits of agreement. The smaller the 95%LOA and closer the bias to the point of equality (Y=0), the more accurate the activity monitor is.

*****Due to association between bias and magnitude of measurement, the Bland-Altman regression method was used to determine mean bias as a function of the mean count [(test count + gold standard count)/2].

Key: 95% LOA = 95% Limits of Agreement, MAPE = Mean absolute percentage error, LFE = Low frequency extension algorithm.

Activity Monitor	Two minute walk test (n=27)			Stairs (n=26)		
	Mean bias (95%CI)	95% LOA	MAPE % (SD)	Bias (95% LOA)	95% LOA	MAPE % (SD)
***Fitbit-zip* (waist)**	10 (-3, 23)	-55, 74	6 (17)	1 (-1, 3)	-8, 10	12 (20)
***Fitbit-zip* (shirt)**	12 (-2, 27)	-58, 83	6 (19)	-6 (-22, 9)	-81, 68	64 (258)
***Actigraph* (-LFE)**	141+(-0.5*mean count)*	110+(-0.5*mean count)*	30 (22)	4 (2, 6)	-4, 12	22 (16)
***Actigraph* (+LFE)**	20 (8, 33)	-40, 81	10 (16)	0 (-1, 1)	-5, 5	8 (9)

The table illustrates the key statistics taken from Bland-Altman analysis. Bias indicates the mean difference between the gold standard step-count and device, whilst the 95%LOA demonstrates the upper and lower limits of agreement. The smaller the 95%LOA and closer the bias to the point of equality (Y=0), the more accurate the activity monitor is.

*****Due to association between bias and magnitude of measurement, the Bland-Altman regression method was used to determine mean bias as a function of the mean count [(test count + gold standard count)/2]

Key: 95% LOA = 95% Limits of Agreement, LFE = Low frequency extension algorithm, MAPE = Mean absolute percentage error, SD=standard deviation.