


BRIEF REPORT

Comparison of Physical Activity Measures Derived From the Fitbit Flex and the ActiGraph GT3X+ in an Employee Population With Chronic Knee Symptoms

Pamela Semanik,¹  Jungwha Lee,² Christine A. Pellegrini,³ Jing Song,² Dorothy D. Dunlop,² and Rowland W. Chang²

Objective. We examined the accuracy of data from an affordable personal monitor (Fitbit Flex) compared with that of data from a research-grade accelerometer worn simultaneously for 7 days; high accuracy would support substitution with this less-expensive personal activity monitor in future community-based arthritis research.

Methods. Subjects (N = 35) with chronic knee symptoms were recruited for a pilot intervention study using Fitbits to increase physical activity in employees with chronic knee symptoms at an urban corporation. Subjects simultaneously wore for 7 days a Fitbit Flex (wrist-worn) and ActiGraph GT3X+ (waist-worn). Fitbit Flex data were regularly stored on a research storage service (Fitabase) by participants. Bland–Altman plots were constructed to examine the agreement between the mean daily times spent in light activity and in bouts of moderate-to-vigorous physical activity (MVPA). Comparisons were calculated by matching Fitabase data from calendar days the Fitbit was worn with data from valid monitoring days (greater than or equal to 10 hours wear time) of the ActiGraph.

Results. Participants at baseline were mostly female (69%) and white (57%) and had a mean age of 52 years and body mass index of 32 kg/m². Bland–Altman analyses indicated systematic bias overall (the Fitbit overestimated both light-intensity activity and MVPA compared with the ActiGraph). The average error varied in magnitude and direction with changing activity amounts.

Conclusion. The Fitbit Flex does not appear to be an adequate substitute for research-grade accelerometry (which represents the gold standard for objective research monitoring of all physical activity intensity levels) in this population of persons with chronic knee symptoms.

INTRODUCTION

Physical activity (PA) can improve strength and function and decrease pain in persons with arthritis, but among adults with lower-extremity joint conditions, as many as four in five do not attain recommended PA thresholds (1). There is increasing public health interest in the objective measurement of PA in free-living persons with arthritis, but gold standard research-grade wearable monitors, such as the ActiGraph GT3X+, can be prohibitively expensive for large-scale studies. The lower cost and availability of

commercial consumer monitors would allow for larger population studies if accuracy is comparable. In addition, consumer monitors' biofeedback features hold an appeal for developing interventions to improve PA as well as improving participants' compliance to the wearable technology, heightening interest in the accuracy of the PA data generated by them.

The general public's broad interest in tracking personal PA with increasingly advanced wearable technology presents an opportunity to capitalize on personal monitor ownership to estimate PA habits of the arthritis population. Attempts to validate

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Supported in part by the National Institute for Arthritis and Musculoskeletal Diseases (grants 1R21-AR-065054-01A1, P60-AR-064464, and 1P30-AR-072579-01) as well as the NIH National Center for Advancing Translational Sciences (grant UL1-TR-001422).

¹Pamela Semanik, PhD: Rush University College of Nursing, Chicago, Illinois; ²Jungwha Lee, PhD, MPH, Jing Song, MS, Dorothy D. Dunlop, PhD, Rowland W. Chang, MD, MPH: Northwestern University Feinberg School of

Medicine, Chicago, Illinois; ³Christine A. Pellegrini, PhD: University of South Carolina School of Public Health, Columbia.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Pamela Semanik, PhD, Rush University College of Nursing, Department of Adult Health and Gerontological Nursing, 600 South Paulina Street, 1080AAC, Chicago, IL 60612. E-mail: Pamela_A_Semanik@rush.edu.

Submitted for publication October 1, 2019; accepted in revised form October 4, 2019.

SIGNIFICANCE & INNOVATION

- This study provides evidence that substitution of research-grade accelerometers with an affordable personal monitor (Fitbit Flex) in community-based studies of physical activity in persons with chronic knee symptoms is not supported.
- This study provides new information on the accuracy of data from an affordable personal monitor (Fitbit Flex) compared with data from a research-grade accelerometer worn simultaneously for 7 days in community-dwelling persons with knee symptoms.

existing consumer-wearable technology (eg, Fitbit) with research-grade monitors (eg, ActiGraph) for agreement in activity time allocated to specific PA intensity levels have yielded mixed results in samples of persons without arthritis (2–4). There has been no known examination of persons experiencing chronic knee symptoms. Strong agreement between the affordable personal monitors and the expensive research-grade accelerometers would support substitution with personal monitors in future PA studies of persons with arthritis.

Although consumer-grade PA monitors have been examined with healthy volunteers walking at speeds and cadences relevant to those of clinical rehabilitation populations (5), it is reasonable to question the validity of such PA monitors in persons with knee symptoms and possible gait alterations other than speed and cadence. Persons with knee symptoms may move more slowly and with significant alterations in hip-, knee-, and ankle-joint function during gait (6) or may adapt their gait to reduce knee joint loads to decrease pain flares (7). It is unknown how these variabilities may affect the accuracy of personal PA trackers. Therefore, we sought to examine the accuracy of the Fitbit Flex in measuring PA in adults with knee symptoms over a 7-day period. Specifically, the time spent in light-, moderate-, and vigorous-intensity activities was compared between the Fitbit Flex and the ActiGraph GT3X+ accelerometer.

PATIENTS AND METHODS

Participants and methods. This research protocol was approved by the university institutional review board. All participants gave informed consent. Employees were recruited for a three-arm randomized clinical trial of the pilot intervention, MobilWise, in which a remote coach viewed PA data generated by a Fitbit personal monitor and used that data to formulate and provide tailored behavioral support using motivational interviewing. The groups included the MobilWise ($n = 19$), Fitbit Only ($n = 16$), and Waitlist Control groups ($n = 16$) at an urban insurance company. Recruitment occurred via a customized website, the link for which was disseminated in corporate announcements. This website detailed the study

requirements then directed interested employees to an initial online screening questionnaire and consent form. The recruitment material messaging was tailored to attract employees with chronic knee symptoms who wanted to increase their PA.

Inclusion criteria for the parent study. To be eligible for this study, employees needed to work full- or part-time for the Chicago office of this company. Whereas most employees commuted to the downtown office at least 4 d/wk, five participants worked primarily from home. Participants had to be older than 18 years of age, have chronic knee symptoms, be able to ambulate at least 15.24 m, be able to speak and read English, and have a body mass index (BMI) less than 40 kg/m².

Exclusion criteria. Potential participants were excluded if an increase in PA was contraindicated by a comorbid condition (screening instruments [Physical Activity Readiness Questionnaire (PAR-Q)] were reviewed and followed up with an interview and/or physical examination by the principal investigator when indicated to assure participant safety), if a total joint replacement had occurred or was planned within the year, if fibromyalgia or inflammatory arthritis was a primary diagnosis, or if the potential participant had a comorbidity that was more functionally limiting than the knee symptoms (eg, spinal stenosis, peripheral vascular disease, or residual effects of stroke). After informed consent was obtained, participants were further screened in person for height and weight (BMI) as well as for the presence of the following:

- uncontrolled diabetes (hemoglobin A1c value greater than 9);
- uncontrolled hypertension (systolic blood pressure level greater than 160 mm Hg or diastolic blood pressure level greater than 110 mm Hg); and
- cardiac risk by history (PAR-Q).

Inclusion criteria for this substudy. Participants had to have been randomized to one of the parent study's two PA promotion intervention arms: MobilWise or Fitbit Only. Data from all participants active at 3 months ($N = 35$) was used.

Measurement. As part of a follow-up evaluation after week 12 of the two pilot intervention groups, subjects simultaneously wore a Fitbit Flex (wrist-worn) and ActiGraph GT3X+ (waist-worn) for 7 days except during water sports or bathing. Participants were encouraged to wear the Fitbit Flex 24 h/d, but the ActiGraph GT3X+ instructions directed participants to wear the unit during waking hours only. Fitbit data were accessed and downloaded from Fitabase and then stored on the secure university server for analyses. The ActiGraph GT3X+ units were collected in person at the work site; accelerometer data were visually inspected for completeness and then stored on the same secure university server. Average daily PA measures were computed for each participant ($N = 35$). Overall, participants generated 226 valid days of moni-

Table 1. Comparison of average daily PA measurements from the Fitbit Flex and ActiGraph GT3X+ worn simultaneously (N = 35 persons)

PA Intensity (min/d)	Fitbit Flex Obtained Data, Median (IQR)	ActiGraph GT3X+ Obtained Data, Median (IQR)	Median Difference, Fitbit – ActiGraph (IQR)	Spearman Correlation (95% CI)
Light	180.4 (137.9 to 251.7)	236.6 (189.1 to 286.3)	–28.3 (–87.3 to –2.7)	0.60 (0.34 to 0.78)
Moderate (bouted)	10.6 (5.6 to 24.6)	10.6 (3.6 to 25.7)	–0.1 (–8.1 to 6.0)	0.52 (0.22 to 0.73)
Vigorous (bouted)	11.6 (6.3 to 27.7)	0 (0 to 0)	11.6 (6.3 to 27.7)	0.25 (–0.09 to 0.54)
MVPA (bouted)	25.0 (13.2 to 62.6)	12.0 (3.7 to 25.7)	11.0 (4.8 to 31.3)	0.73 (0.52 to 0.85)

Abbreviation: CI, confidence interval; IQR, interquartile range; MVPA, moderate-to-vigorous physical activity; PA, physical activity.

toring (a valid monitoring day was defined as greater than or equal to 10 h/d of wear time).

The parameters of PA-intensity categories were defined for each measurement device. The thresholds for the proprietary Fitbit categories were based on metabolic equivalent task (MET) calculations detailed by Fitabase (E. Ramirez, PhD, May 2018, personal written communication). The “lightly active” Fitbit category included activity registering between 1.5 and 3 METs. The “fairly active” category included activity registering between 3 and 6 METs in at least 10-minute bouts. The “very active” Fitbit category included activity registering at greater than or equal to 6 METs or greater than or equal to 145 steps per minute in at least 10-minute bouts. Lastly, the “active” Fitbit category (fairly active + very active = a minimum of 3 METs or more in at least 10-minute bouts) comprised what is generally considered moderate-to-vigorous physical activity (MVPA).

Following convention, the National Institutes of Health (NIH) accelerometer thresholds for activity intensity were used for defining ActiGraph activity categories based on vertical counts per minute (8). Light activity was defined as 100 to 2019 cpm, moderate activity was defined as 2020 to 5998 cpm in at least 10-minute bouts, vigorous activity was defined as 5999 cpm and more in at least 10-minute bouts, and MVPA was defined as 2020 cpm and more in at least 10-minute bouts. Bouted minutes were calculated with allowance for interruptions of 1 or 2 minutes below the thresholds.

Analyses. Data from the days that the Fitbit was worn were compared with data from days that the ActiGraph GT3X+ was worn (valid monitoring days were defined as 10 h/d or more of wear time). Histograms of all data were constructed and inspected. A correlation table (Table 1) was constructed to examine the associations between the average daily amount of time spent in individual activity-intensity categories (light, moderate, vigorous, and MVPA; the last 3 categories in bouts of 10-minutes or more).

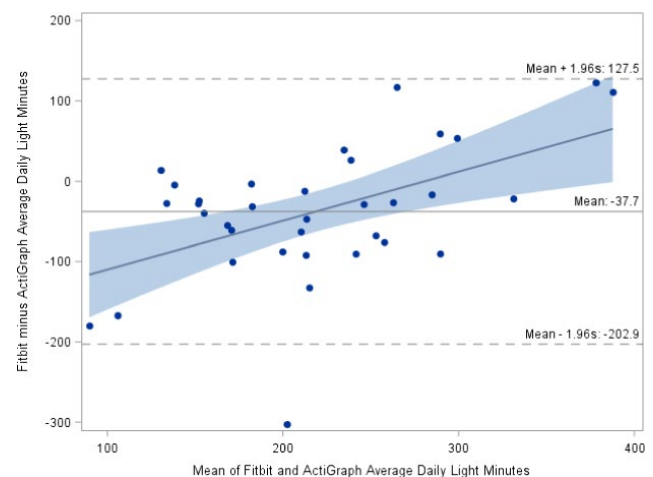
Bland–Altman plots were used to visualize any systematic differences between the two highest correlations: average daily light-activity time ($p = 0.60$; 95% confidence interval [CI]: 0.34–0.78) and bouted MVPA time ($p = 0.73$; 95% CI: 0.52–0.85) from the two measurement devices. The differences between the Fitbit and ActiGraph GT3X+ estimates (y-axis) were plotted against the means of the estimates from the two devices (x-axis) for light

activity and bouted MVPA. The regression line of the difference (with 95% confidence limits) was plotted to detect proportional differences along with 95% limits of agreement (mean difference $\pm 1.96 \times$ SD of the differences) for visual examination to evaluate the global agreement between the measurements from the two devices. A horizontal line at zero would represent complete agreement and no bias. Data were analyzed using SAS version 9.4 (SAS Institute).

RESULTS

Participants (N = 35) were mostly female (69%) and white (57%) and had a mean age of 52 years and a mean BMI of 32 kg/m².

To examine the data from the two devices for potential bias and direction of bias, Bland–Altman plots were used to compare the agreement between the Fitbit and ActiGraph GT3X+ estimates of both light activity (Figure 1) and bouted MVPA (Figure 2). These strongly sloping regression lines not only show that Fitbit measures are biased when compared with ActiGraph GT3X+ measures but also show that the difference in measures increases with greater amounts of light activity or bouted MVPA. Most of the differences lie between 95% limits of agreement for light-intensity PA; however, the SDs of the differences (SD = 84.3) are quite large compared with the mean differences.

**Figure 1.** Bland–Altman plot of Fitbit versus ActiGraph daily light-intensity physical activity from N = 35 participants.

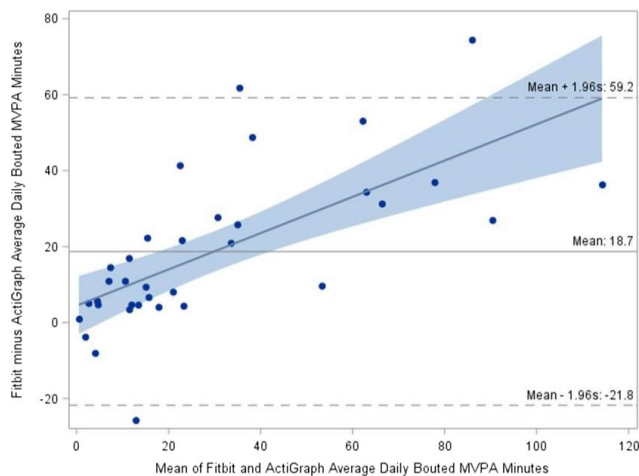


Figure 2. Bland–Altman plot of Fitbit versus ActiGraph daily bouts moderate-to-vigorous physical activity (MVPA) from $N = 35$ participants.

As shown in Figure 1, the Fitbit underestimated light-activity minutes compared with the ActiGraph GT3X+ at times of relatively low activity amounts but overestimated light-activity minutes as light-activity amounts increased. The amount of under- or overestimation varied by the number of minutes of light activity.

In the Figure 2 Bland–Altman plot, bouts MVPA minutes are evaluated. On average, there is a 20-minute bias, but bias is not consistent. The Fitbit overestimated MVPA compared with ActiGraph GT3X+, but the amount of overestimation increased as the number of minutes of MVPA increased. Although most of the points are within the limits-of-agreement lines, the limits of agreement are very wide.

DISCUSSION

To our knowledge, this was the first attempt to validate existing consumer-wearable technology with research-grade PA monitors in persons with chronic knee symptoms. Using data collected entirely in a free-living sample of primarily middle-aged white women with overweight, the Fitbit registered less activity than the ActiGraph GT3X+ in the lower-PA-intensity ranges and registered more activity than the ActiGraph GT3X+ at the higher-intensity ranges. Bland–Altman plots showed systematic bias in measures of both light-intensity activity and MVPA, but that bias varied as the number of minutes in each activity-intensity category increased. Thus, there does not appear to be a way to correct for these discrepancies.

Bland–Altman analyses findings from other study populations in which data from these two devices were compared have varied. In comparisons of minutes spent in MVPA, results have varied. Sushames et al (2) found that in healthy adults, the average MVPA minutes measured by the Fitbit Flex was significantly lower compared with that measured by the ActiGraph. However, it appears that in their evaluation, they compared

ActiGraph total MVPA (not bouts) with Fitbit (bouts) MVPA, which may account for that difference. According to the Fitbit website, all reported MVPA is bouts by default (9). Conversely, when Dominick et al (3) compared minute-level data from both the Fitbit Flex and ActiGraph in healthy young adults, the Fitbit significantly underestimated the proportion of time in light-intensity activity by 34% and overestimated by 3% time spent in both moderate- and vigorous-intensity activity (all $P < 0.001$). Most recently, researchers testing the validity of the Fitbit Flex compared with the ActiGraph GT3X+ in younger healthy participants also found evidence of systematic bias in their Bland–Altman analyses, indicating that the Fitbit Flex overestimated mean daily MVPA. They also noted that the slope for the fit line suggested that the discrepancy tended to increase as the total mean daily MVPA volume increased (10).

Our results were compared with those from two systematic reviews of validity and reliability of consumer-wearable activity trackers, which included Fitbits. In their review, Evenson et al (11) did not focus primarily on the amount of time spent in PA-intensity categories. The review did include two studies of the Fitbit Zip model, which either correlated well with accelerometer readings or generally overcounted minutes of MVPA (11). In their review, Feehan et al (12) focused on the accuracy of measures derived from Fitbit devices and noted that there was a tendency for the Fitbit to overestimate MVPA in free-living settings compared with an ActiGraph accelerometer, similar to our study.

The proprietary Fitbit algorithms for calculating time spent within PA-intensity categories and how these algorithms may have changed over time is not known. It may be that Fitbit algorithms are geared toward detecting bouts higher-intensity activity, which was favored by the US PA guidelines during the time these participants wore the devices (10). As Gomersall et al (13) have pointed out, “increased transparency from manufacturers regarding exact definitions of their variables and how they are calculated (including both idle time and active time...) would significantly improve the ability of researchers to explore the accuracy of these devices.” However, appealing to researchers (as opposed to consumers) may not be the industry’s goal.

Given these differences, it does not appear that the Fitbit Flex is an adequate substitute for research-grade accelerometry in endeavors to compare PA in populations. This does not preclude the usefulness of the Fitbit to provide participant feedback on PA in intervention studies. However, feedback from these commercial devices should be interpreted with caution. If commercial-grade devices do indeed overestimate MVPA, what may be a modest discrepancy on any given day can lead to gross misconception about meeting PA guidelines over the course of the week. Use of the device to gauge improvement in activity levels over time, as opposed to absolute levels of PA within an intensity category, may be the best use.

The generalizability of findings is limited because of the predominantly female, middle-aged sample with knee symptoms. The

study could not control for wear time of the consumer devices, and this might have impacted results. Some differences in activity time may be related to the Fitbit possibly being worn 24 hours (versus the waking hours that participants were instructed to wear the ActiGraph GT3X+). One potential confounder includes the wrist versus waist location of the monitoring devices during data collection, although this arrangement is consistent with that of similar studies that compared data from these devices (2–4,10,13–16). Because it is not known how the Fitbit analyzes data from its devices, it is possible that the differences that we noted may be due to the Fitbit using vector magnitude in its data processing or a different epoch length in its algorithm for its calculations (eg, 30 seconds versus the 60-second epoch length we used to analyze ActiGraph data). However, others have noted in sensitivity analyses that data processing with alternate epoch lengths for ActiGraph data in comparison with Fitbit data did not alter the overall findings (13).

This comparison of PA data derived from the Fitbit Flex and ActiGraph GT3X+ not only found systematic bias but also found that the magnitude and direction of the average device error changed as the number of minutes in each activity-intensity category increased. Based on these findings, the Fitbit Flex does not appear to be an adequate substitute for research-grade accelerometry, which represents the gold standard for objective research monitoring of all PA intensity levels in this population of persons with chronic knee symptoms.

ACKNOWLEDGMENTS

Janie Urbanic, MA, LPC, Rush College of Nursing; Corporate Wellness Team; Fitbit Corporation.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Semanik, Lee, Pellegrini, Song, Dunlop, Chang.

Acquisition of data. Semanik, Pellegrini, Song.

Analysis and interpretation of data. Semanik, Lee, Pellegrini, Song, Dunlop, Chang.

REFERENCES

- Dunlop DD, Song J, Semanik PA, Chang RW, Sharma L, Bathon JM, et al. Objective physical activity measurement in the osteoarthritis initiative: are guidelines being met? [Original Data Report]. *Arthritis Rheum* 2011;63:3372–82.
- Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One* 2016;11:e0161224.
- Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. *JMIR Mhealth Uhealth* 2016;4:e110.
- Alharbi M, Bauman A, Neubeck L, Gallagher R. 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;23:1476–85.
- Singh AK, Farmer C, van den Berg ML, Killington M, Barr CJ. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. *Disabil Health J* 2016;9:320–3.
- Farrokhi S, O'Connell M, Fitzgerald GK. Altered gait biomechanics and increased knee-specific impairments in patients with coexisting tibiofemoral and patellofemoral osteoarthritis. *Gait Posture* 2015;41:81–5.
- Boyer KA, Hafer JF. Gait mechanics contribute to exercise induced pain flares in knee osteoarthritis. *BMC Musculoskelet Disord* 2019;20:107.
- Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;40:181–8.
- Fitbit Inc. What are active minutes? Updated April 25, 2019. URL: https://help.fitbit.com/articles/en_US/Help_article/1379.
- Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC Med Res Methodol* 2019;19:29.
- Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;12:159.
- Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth Uhealth* 2018;6:e10527.
- Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. *J Med Internet Res* 2016;18:e239.
- Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLoS One* 2017;12:e0172535.
- Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. *Br J Sports Med* 2018;52:844–50.
- Chow JJ, Thom JM, Weweg MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. *Gait Posture* 2017;57:199–203.