Validating Fitbit for Evaluation of Physical Activity in Patients with Knee Osteoarthritis: Do Thresholds Matter?

Genevieve S. Silva,1 Heidi Yang,1 Jamie E. Collins,2 and Elena Losina2

INTRODUCTION

Suboptimal physical activity (PA) is a health concern around the world (1), which has led public health organizations to release official PA guidelines (2,3). In 2008, the US Department of Health and Human Services published suggested weekly PA guidelines for Americans (2), stratified by intensity of activity. These guidelines were updated in 2018 to include strategies for population-level PA promotion, emphasize the risks of sedentary behavior, and eliminate the previous requirement that PA occur in bouts of at least 10 minutes (4). In conjunction with this emphasis on PA guidelines, studies have attempted to quantify PA levels currently achieved in different population groups (5–7).

Patients with symptomatic knee osteoarthritis (OA) have suboptimal levels of PA (8–12). PA uptake among knee OA patients remains low despite growing evidence that PA and exercise are efficacious for pain relief (13). These observations highlight the importance of developing and implementing interventions to improve PA among knee OA patients.

To evaluate the efficacy of such interventions, it is important to accurately measure PA levels. Triaxial, medical-grade accelerometers like the ActiGraph GT3X+ (ActiGraph Corporation) are considered the gold standard for PA evaluation (14), yet guidelines for interpreting accelerometry data are not standardized in the literature (15). Studies have proposed multiple threshold cutoffs to relate ActiGraph counts to PA levels, and some have

Objective. We sought to evaluate the performance of Fitbit in estimating ActiGraph-derived moderate-to-vigorous physical activity (MVPA) and sedentary time in the knee osteoarthritis (OA) population.

Methods. We used data from two weeks of Fitbit and ActiGraph wear among knee OA subjects. In primary analyses, we used literature-based ActiGraph thresholds of 200 and 1924 counts/min (triaxial vector magnitude) for sedentary and MVPA as the gold standard to which we compared three sets of Fitbit thresholds informed by literature and data (Youden index). We also considered personalized, stride length–based Fitbit thresholds. In sensitivity analyses, we used uniaxial, vertical axis–based as well as personalized, BMI-based ActiGraph thresholds. We calculated agreement, sensitivity, and specificity of Fitbit in classifying sedentary and MVPA time.

Results. In the primary analysis (vector magnitude thresholds), maximum agreement for sedentary and MVPA time was 67.0% from the Youden index–based and 91.1% from the stride length–based Fitbit thresholds. For sedentary time, the 20 strides/min threshold had the highest sensitivity (97.6%), and Youden-derived 1 stride/min had the highest specificity (51.6%). For MVPA, Youden-derived 14 strides/min yielded 72.8% sensitivity, and using stride length yielded 98.6% specificity. MVPA time ranged from 49–323 min/d, depending on threshold used, with literature-based and personalized thresholds leading to more conservative estimates of MVPA than Youden-derived thresholds.

Conclusion. Using Fitbit for MVPA and sedentary time assessment may lead to inaccurate estimates of both. Fitbit MVPA estimates were generally more conservative than ActiGraph estimates. Incorporating individuals’ characteristics did not meaningfully improve Fitbit performance. Caution should be exercised when measuring activity using Fitbit.

ACR Open Rheumatology
Vol. 1, No. 9, November 2019, pp 585–592
DOI 10.1002/acr2.11080
© 2019 The Authors. ACR Open Rheumatology published by Wiley Periodicals, Inc. on behalf of American College of Rheumatology. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1Genevieve S. Silva, BS, Heidi Yang, MS: Brigham and Women’s Hospital, Boston, Massachusetts; 2Jamie E. Collins, PhD, Elena Losina, PhD: Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts.

Dr. Collins received funding from Genentech (to BWH) and is a consultant to Boston Imaging Core Labs (<$10,000). Dr. Losina received funding from Pfizer, Samumed, Flexion, and Genentech (funding to BWH). She serves as an advisor to Velocity (<$10,000). No other disclosures relevant to this article were reported.

Address correspondence to Elena Losina, PhD, Orthopaedic and Arthritis Center for Outcomes Research (OrACoRe), Policy and Innovation Evaluation in Orthopaedic Treatments (PIVOT) Center, Department of Orthopaedic Surgery, Brigham and Women’s Hospital, 75 Francis Street, BET 5-016, Boston, MA 02115. E-mail: elosina@bwh.harvard.edu.

Submitted for publication March 28, 2019; accepted in revised form August 20, 2019.
suggested that thresholds should be described as ranges to account for interperson variability (15–19). Furthermore, much accelerometry validation has involved young and healthy populations or uniaxial ActiGraphs (15,20–22), the results of which are not applicable to the older OA population using triaxial accelerometers (23).

Although ActiGraphs are the gold standard for objective PA assessment, commercially available accelerometers like the Fitbit have distinct advantages for use in PA-focused investigations. Such devices have a significantly lower cost and greater ease of wear, factors that make them appealing for scalability in large trials conducted over extended periods of time, leading to their increased use in research (24–28). Although studies have begun to investigate agreement between medical- and commercial-grade accelerometers (29–32), standardized thresholds across devices have not been delineated, particularly those specific to at-risk populations with unique gait patterns, such as older individuals with knee OA (33). As Fitbit use becomes more commonplace in large-scale investigations of the OA population, it is essential to understand the meaning of the data in the context of validated ActiGraph accelerometry measures.

We sought to examine a variety of thresholds—personalized, data-driven, and published—to determine the accuracy of the Fitbit Charge 2 in approximating moderate-to-vigorous physical activity (MVPA) and sedentary time, compared with the gold standard ActiGraph GT3X+, under various classification paradigms. Our primary purpose in this analysis was to examine agreement between the two accelerometers in measuring activity level. Therefore, we only considered time during which both devices registered wear, per our definitions. Wherever possible, we focused on thresholds applicable to an older adult population representative of individuals most affected by knee OA.

METHODS

Study sample. We analyzed data collected in a pilot study of 15 participants with symptomatic knee OA recruited from the patient population of four rheumatologists at Brigham and Women’s Hospital (34). Participants were recruited between March and July of 2017. Participants were asked to simultaneously wear Fitbit (wrist) and ActiGraph GTX3 (hip) accelerometers over the course of 2 weeks for at least 10 waking hours per day, with which 14 of the 15 participants complied. In our previous report (34), we established that these sites (wrist for Fitbit and hip for ActiGraph) offered the best agreement across the accelerometers. Fitbit data were downloaded from the Fitbit server on a weekly basis using an in-house application programming interface, and ActiGraph data were downloaded at the end of the study period using ActiLife software (ActiLife v.6.13.3, ActiGraph). Further details of this pilot study have been published elsewhere (34).

Wear time description and validation. To validate Fitbit wear time at the per-minute level, we compared Fitbit stride counts to heart rate (HR) readings. We examined each minute of data for each of the 14 study participants included, classifying HR values as either missing or greater than zero and stride counts as equal to zero or greater than zero to define wear time and sedentary time without using stride counts alone.

In all analyses, we defined nonwear time for the ActiGraph as any minute with zero vector magnitude counts. We defined a Fitbit nonwear minute as having a stride count of zero along with missing HR data. Fitbit minutes with zero strides and nonmissing HR data were considered as sedentary time. Any Fitbit minute with greater than zero strides was considered as wear time, even if HR data were missing. Our analysis included minutes during which both the ActiGraph and Fitbit were worn.

Defining Fitbit thresholds: main analysis. We compared Fitbit to ActiGraph data for each subject on a minute-by-minute level to examine agreement between the devices. Minutes from each device were classified to represent time in one of two activity categories: sedentary or MVPA. We determined these categories by defining stride or count (for Fitbit and ActiGraph, respectively) threshold cut point values for the upper bound of sedentary activity and the lower bound of MVPA.

We considered four threshold sets for Fitbit-derived PA data. We included two combinations of upper and lower strides/min cut points, which were informed by literature-suggested speed and average stride length, and we used one data-driven threshold set defined by a Youden index analysis (35). We also calculated a personalized threshold for MVPA by utilizing the stride length data for each individual (34).

A walking speed of 2.5 km/h was considered consistent with MVPA for an older population (16), particularly one with elevated pain (mean Knee Injury and Osteoarthritis Outcome Score [KOOS] pain = 50) and limited function due to OA. We used an estimated stride length of 0.6 m (derived from a measured step length of 0.3 m) to conclude that at least 67 strides/min would be needed to achieve a speed of 2.5 km/h. We considered an alternate MVPA lower bound of 50 strides/min, which corresponded to the ~2 km/h pace suggested by another report for an elderly population (36). We considered two upper thresholds for sedentary time, 10 and 20 strides/min, which corresponded to speeds of about 0.4 and 0.7 km/h, respectively (37).

We also defined a Fitbit threshold with a data-driven approach. Using the ActiGraph as the gold standard, we used the receiver operating characteristic curve to evaluate possible cut points in Fitbit-measured strides. We used the Youden index to select the cut points with the maximum combination of sensitivity and specificity (35). This yielded an upper bound for sedentary time of 1 stride/min and a lower bound for MVPA of 14 strides/min.
To determine individualized Fitbit thresholds, we used coefficients for metabolic equivalents of task (METs) from Barnett et al. and calculated METs for each participant based on their stride lengths recorded by the Fitbit (16). MVPA was defined as expending 3 METs of energy, which, for older adults, corresponds to walking at a 2.5 km/h pace (16). We calculated personalized MVPA thresholds from this by dividing that speed by half of each participant’s stride length (34, 38).

We used the published ActiGraph thresholds of 200 counts/min for the upper bound of sedentary time (18) and 1924 counts/min for the lower bound of MVPA, which were derived for an elderly population as the gold standard (16).

We then compared the overall agreement, sensitivity, and specificity of the Fitbit compared with the ActiGraph in classifying minutes spent in sedentary and MVPA time. Agreement for each PA level was calculated by summing the overall percentage of minutes classified by both devices as the same PA level. Sensitivity represented the ability of the Fitbit to correctly identify sedentary and MVPA time, and specificity represented the ability of the Fitbit to correctly identify time not in sedentary time and not in MVPA, relative to the ActiGraph classifications.

Sensitivity analyses: varying ActiGraph threshold definitions. Uniaxial ActiGraph thresholds. As a sensitivity analysis, we repeated the above analyses using uniaxial ActiGraph thresholds as the gold standard against which the same set of Fitbit results were compared. For the uniaxial counts, the ActiGraph cutoffs were 100 counts/min as the upper bound of sedentary time (11) and 1013 counts/min as the lower bound of MVPA time (16). For the uniaxial ActiGraph data, we considered any minute with a vertical (uniaxial) count of zero as sedentary time if that same minute had a vector (triaxial) count greater than zero, indicating wear. Otherwise, uniaxial counts of zero were classified as non-wear.

Per-person ActiGraph thresholds by body mass index adjustment. Additionally, some literature has suggested that the thresholds for determining MVPA may differ by body mass index (BMI) (16). To investigate this, we used the regression coefficients provided by Barnett et al. to adjust our ActiGraph count thresholds by BMI for each subject (16). We compared Fitbit data, using the thresholds defined in our primary analysis, to ActiGraph hip minute-by-minute data with MVPA thresholds adjusted by BMI for each subject. We again examined agreement, sensitivity, and specificity between the devices, and we did this for both triaxial and uniaxial ActiGraph counts to evaluate MVPA versus non-MVPA time.

Estimating PA. We also compared the average daily minutes classified as MVPA time across our Fitbit MVPA thresholds (50, 67, Youden-derived [14 with triaxial and 15 with uniaxial], and stride length–derived strides/min) and our unadjusted (1924 triaxial and 1013 uniaxial counts/min) and BMI-adjusted MVPA thresholds for triaxial and uniaxial ActiGraph counts.

RESULTS

Description of the study sample. Table 1 presents the cohort characteristics of the study participants, who had a mean(SD) stride length of 0.6(0.1) and a mean(SD) BMI of 28.9(5.4). Examining Fitbit wear time independently of ActiGraph wear time, subjects wore the Fitbit (defined as all minutes except those with both 0 strides and missing HR) for a total of 2217 hours (133011 minutes) over the course of 2 weeks, or approximately 11 hours per subject per day. Seventy-four percent of those minutes did not register any strides, whereas 26% (33657 minutes) registered at least one stride. Less than 0.5% of the minute-by-minute Fitbit data represented instances in which stride count was greater than zero and HR was missing.

The total time that subjects wore both the Fitbit and triaxial ActiGraph was 93087 minutes, and the mean per-subject wear time classified by both Fitbit and ActiGraph was 6649 minutes (111 h/subject), or approximately 8 hours per subject per day.

Fitbit PA categorization versus triaxial ActiGraph gold standard, varying Fitbit thresholds. Results of all Fitbit cut point set scenarios (three sets for sedentary time and four sets for MVPA) compared with the triaxial ActiGraph hip are reported in Table 2. The Youden analysis resulted in an optimal Fitbit cutoff of 1 stride/min as the upper bound of sedentary and 14 stride/min for the lower bound of MVPA. The stride length–based MVPA Fitbit threshold ranged from 57-87 strides/min, with a mean(SD) of 70(9) strides/min.

Among the scenarios evaluated in the primary analysis, the Fitbit threshold that yielded the greatest agreement with the ActiGraph for classifying sedentary time was the Youden-derived threshold of 1 stride/min (67.0%), with the stride length–derived Fitbit threshold yielding greatest agreement (91.1%) for MVPA classification (Table 2). Overall agreement for sedentary and MVPA time ranged from 55.5% to 67.0% and from 81.6% to 90.8%, respectively.

Table 1. Description of the study sample (n = 14)

| Characteristic          | Value* |
|-------------------------|--------|
| Age (years)             | 68.0 (8.3), 68.0 |
| Female sex, n (%)       | 9 (64.3) |
| BMI (kg/m²)             | 28.9 (5.4), 27.7 |
| Stride length (m)       | 0.6 (0.1), 0.6 |
| Medication for knee pain in past week, n (%) | 12 (85.7) |
| KOOS Scores*            |        |
| Pain                    | 49.6 (20.9), 56.7 |
| Function                | 55.3 (20.5), 60.7 |
| Symptoms                | 50.2 (18.1), 50.0 |
| Quality of life         | 38.2 (17.8), 42.5 |
| Triaxial ActiGraph and Fitbit wear time (minutes) | 93087 |

Abbreviation: BMI, body mass index; KOOS, Knee Injury and Osteoarthritis Outcome Score.

*Values are mean(SD), median unless otherwise specified. *KOOS scores: 0-100; 0 represents the worst outcome score (42).
Table 2. Fitbit specificity, sensitivity, and agreement with ActiGraph in identifying minutes in sedentary and MVPA time based on various threshold definitions

| Fitbit-based threshold (strides/min) | Sedentary Time | MVPA Time |
|-------------------------------------|----------------|-----------|
|                                     | Sensitivity (%) | Specificity (%) | Agreement (%) | Sensitivity (%) | Specificity (%) | Agreement (%) |
| **Triaxial ActiGraph; Non-BMI**     |                |            |               |                |            |               |
| (sedentary threshold: 200, MVPA threshold: 1924) |                |            |               |                |            |               |
| 10                                  | 94.9           | 41.2       | 62.3          | 50              | 36.0        | 97.0          | 90.8          |
| 20                                  | 97.6           | 28.4       | 55.5          | 67              | 23.5        | 98.5          | 90.9          |
| 1†                                 | 90.8           | 51.6       | 67.0          | 14†             | 72.8        | 82.6          | 81.6          |
|                                     |                |            |               | **Stride length** |            |            |               |
|                                     |                |            |               | **Uniaxial ActiGraph; Non-BMI** |            |            |               |
| (sedentary threshold: 100, MVPA threshold: 1013) |                |            |               |                |            |            |               |
| 10                                  | 92.5           | 47.9       | 70.9          | 50              | 40.2        | 96.5          | 92.1          |
| 20                                  | 96.7           | 34.1       | 66.4          | 67              | 27.4        | 98.2          | 92.7          |
| 1†                                 | 86.8           | 58.2       | 73.0          | 15†             | 71.9        | 82.0          | 81.3          |
|                                     |                |            |               | **Stride length** |            |            |               |
|                                     |                |            |               | **Triaxial ActiGraph; BMI-Based** |            |            |               |
| (mean MVPA threshold: 1807)         |                |            |               |                |            |            |               |
| 50                                  | 30.5           | 96.8       | 89.2          |                |            |            |               |
| 67                                  | 19.6           | 98.3       | 93.9          |                |            |            |               |
| 9†                                 | 74.6           | 77.5       | 77.2          |                |            |            |               |
|                                     |                |            |               | **Stride length** |            |            |               |
|                                     |                |            |               | **Uniaxial ActiGraph; BMI-Based** |            |            |               |
| (mean MVPA threshold: 903)          |                |            |               |                |            |            |               |
| 50                                  | 26.0           | 96.2       | 88.1          |                |            |            |               |
| 67                                  | 16.8           | 98.0       | 88.5          |                |            |            |               |
| 12†                                | 61.9           | 80.0       | 77.9          |                |            |            |               |
|                                     |                |            |               | **Stride length** |            |            |               |
|                                     |                |            |               |                  |            |            |               |
| Abbreviation: BMI, body mass index; MVPA, moderate-to-vigorous physical activity. |
| †Representation of a threshold derived using the Youden index analysis; *representation of thresholds derived using individualized stride length; The triaxial cutoffs used for sedentary and MVPA time for the ActiGraph were 200 and 1924 counts/minute, respectively. The uniaxial cutoffs were 100 and 1013 counts/minute for sedentary and MVPA time, respectively. BMI-based analyses involved adjusting thresholds on a per-subject basis by BMI calibration, for which data was only available for an MVPA, not sedentary, time threshold. Sensitivity represents the ability of the Fitbit to correctly identify time in MVPA, specificity represents ability of the Fitbit to correctly identify time not in MVPA, and agreement represents the total proportion of minutes categorized correctly by the Fitbit, all relative to the ActiGraph categorizations. |

Among the Fitbit-based thresholds for sedentary time, 20 strides/min had the highest sensitivity of 97.6%. In other words, 97.6% of all minutes designated as sedentary time by the ActiGraph had <20 strides/min measured by the Fitbit, thereby also being classified as sedentary time. Specificity for sedentary time was highest using the Youden-defined 1 stride/min threshold, reaching 51.6%; that is, 51.6% of all minutes designated as nonsedentary time by the ActiGraph had >1 stride/min measured by the Fitbit also being classified as nonsedentary time. For MVPA classification, the Youden-derived 14 strides/min threshold yielded the greatest sensitivity (72.8), and stride length yielded the greatest specificity (98.6).

Figure 1 visualizes the distribution of Fitbit-based strides/min versus ActiGraph-based counts/min. Quadrants 2 (Q2) and 3 (Q3) of Figure 1 represent agreement between the two accelerometers when using 1924 counts/min and 67 strides/min as the lower bounds of MVPA. Although most of the data points are distributed across Q2 and Q3, a nontrivial number covers Q1 and Q4, pointing to the lack of perfect correspondence between MVPA thresholds defined by ActiGraph and Fitbit. The same discrepancy was present across all alternative Fitbit thresholds.

Sensitivity analyses: varying ActiGraph threshold definitions. The results of the sensitivity analyses using uniaxial ActiGraph thresholds and BMI-adjusted thresholds for both triaxial and uniaxial ActiGraph data are also presented in Table 2. BMI-derived MVPA thresholds for the triaxial ActiGraph ranged from 902 to 2456 counts/min, with a mean(SD) of 1807(441) compared with the singular, standardized threshold of 1924 counts/min. The same thresholds for the uniaxial ActiGraph ranged from 88 to 1488 counts/min, with a mean(SD) of 903(397), compared with the singular, standardized MVPA threshold of 1013 counts/min. The unadjusted uniaxial ActiGraph thresholds achieved maximum agreement (73.0%) in identifying sedentary time when paired with the Youden-derived Fitbit threshold of 1 stride/min. Sedentary time sensitivity ranged from 86.8 to 96.7, and specificity ranged from 34.1 to 58.2 (Table 2).

Maximum overall agreement in identifying MVPA time was achieved when using the stride length–based Fitbit threshold when compared with all variations of the ActiGraph thresholds: 93.0% with the unadjusted uniaxial thresholds, 89.5% with the BMI-adjusted triaxial thresholds, and 88.9% with the BMI-adjusted uniaxial thresholds.

Correct classification of MVPA time was maximized when using the Youden-derived Fitbit thresholds for all ActiGraph scenarios, yielding a sensitivity of 71.9, 74.6, and 61.9 for unadjusted uniaxial, BMI-adjusted triaxial, and BMI-adjusted uniaxial ActiGraph-based standards, respectively. Correct classification of time not in MVPA was maximized when using stride length–derived Fitbit thresholds for all ActiGraph scenarios, although the
67 strides/min threshold yielded very similar values (Table 2). The specificity values were 98.4, 98.5, and 98.2, respectively.

**Estimating PA.** Figure 2 depicts the estimated average daily MVPA across the alternative MVPA thresholds and accelerometers. The mean daily MVPA time ranged from 49 to 323 minutes across all Fitbit and ActiGraph threshold scenarios. ActiGraph-based thresholds led to greater daily MVPA time when compared with Fitbit-based thresholds, with the exception of the Youden-driven Fitbit thresholds of 14 and 15 strides/min, which led to the highest estimate of average daily MVPA for triaxial (323 min) and uniaxial (311 min) thresholds, respectively. The difference between MVPA estimates using appropriate uniaxial or triaxial thresholds was negligible.

**DISCUSSION**

We conducted this validation study to determine the accuracy (agreement, sensitivity, and specificity) of alternative Fitbit-based thresholds in identifying MVPA. The results of our analyses suggest that Fitbit-based thresholds have high sensitivity and low specificity for classifying sedentary time and low sensitivity and high specificity for classifying MVPA time. We also found that Fitbit more conservatively estimated MVPA time compared with ActiGraph accelerometers. Fitbit tended to overestimate sedentary time and underestimate MVPA time.

Using literature threshold values of 200 (sedentary) and 1924 (MVPA) counts/min for triaxial ActiGraph accelerometers, the maximum agreement we achieved between Fitbit and ActiGraph classifications of sedentary and MVPA time was 67.0% and 91.1%, respectively, corresponding to Fitbit thresholds derived by the Youden index and stride length. This rather poor agreement for sedentary time illustrates that Fitbit does not estimate sedentary time in alignment with the gold standard when using a single triaxial threshold. However, optimizing MVPA time classification by stride length yielded high agreement with the same gold standard. This participant-adjusted agreement aligns with another study that compared outputs from ActiGraph and Fitbit and found an agreement of classifying time in vigorous PA of $r = 0.80$ (39). In our primary analysis using unadjusted triaxial ActiGraph thresholds, we found that average sensitivity across all Fitbit thresholds was greater for sedentary time than for MVPA time. This corroborates a recent study that found that Fitbit Flex more accurately estimated sedentary time than MVPA time in comparison with the ActiGraph GT3X+ (32).

The data-driven Youden index produced the highest sensitivity and lowest specificity across all ActiGraph threshold scenarios in classifying MVPA. Across all subjects, many more minutes were spent not in MVPA versus in MVPA. The Youden index, which equally weights sensitivity and specificity, had lower overall agreement with the gold standard ActiGraph but was the only threshold to reach >50% sensitivity (identifying >50% of minutes deemed to...
be MVPA by the ActiGraph). Thus, selecting the cut point with the best performance depends on the ultimate goal of the classification exercise.

Agreement, specificity, and sensitivity were higher for all Fitbit scenarios for the unadjusted, literature-based triaxial ActiGraph thresholds than for the BMI-adjusted triaxial ActiGraph thresholds; the same pattern held true for the unadjusted versus BMI-adjusted uniaxial ActiGraph thresholds. Although this does not support previous proposals (16,40) for adjusting accelerometer cut point thresholds based on individual characteristics, we also found that individualized stride length–derived Fitbit thresholds yielded the best MVPA time agreement in our sensitivity analyses. However, these stride length–derived agreement values did not differ greatly from the literature-based agreement values within each ActiGraph threshold scenario. Another study found that ActiGraph underestimated mean steps more than Fitbit did (31). However, we found that the triaxial ActiGraph classified greater average daily minutes in MVPA under both the literature-based and BMI-adjusted threshold scenarios than all Fitbit threshold scenarios, aside from the Youden-derived Fitbit threshold.

We note several limitations and assumptions of the current study. The sample size from which we drew our accelerometry data was a pilot trial with 15 individuals, and only 14 trial participants wore both their Fitbit and ActiGraph during the study period. However, analysis on the minute-by-minute level across 14 days allowed us to analyze a substantial number of data points despite a relatively low number of participants. We defined MVPA using a walking speed (2.5 km/h) below the average speed of healthy adults in order to best represent our study population: older adults (mean age 68) with symptomatic knee OA (mean KOOS pain 50; 0-100 scale with 100 being the worst). In addition, we used the lowest possible “normal” frequency of 30 Hz in our ActiGraph data collection protocol but did not use low-frequency extension (LFE) filters. Although some data show that incorporating LFE increases the ActiGraph’s ability to capture low activity readings, most comparisons used the standard ActiGraph setting of 60 Hz (41). Given that we used a lower frequency setting of 30 Hz in our ActiGraph data collection protocol but did not use low-frequency extension (LFE) filters. Although some data show that incorporating LFE increases the ActiGraph’s ability to capture low activity readings, most comparisons used the standard ActiGraph setting of 60 Hz (41). Given that we used a lower frequency setting of 30 Hz, the impact of using an LFE filter would have been less significant in our context. Finally, as we did not have HR readings from the ActiGraph, we excluded all ActiGraph time with zero counts from this analysis. Therefore, it is likely that we underestimated overall sedentary time.

We conclude that no Fitbit-based thresholds yielded high enough specificity (the highest was 51.6%) to qualify Fitbit as a reasonable proxy for ActiGraph in estimating sedentary time. We found that optimizing Fitbit thresholds by stride length yielded high agreement of MVPA time classification with ActiGraph measures, which is comparable to agreement using literature-based Fitbit thresholds. However, we also recommend
against using Fitbit to determine minutes in MVPA in the older, heavier, knee OA population, as we only achieved sensitivity greater than 50 using the data-driven Youden index to classify Fitbit thresholds. Adjusting ActiGraph MVPA thresholds by participant BMI also does not yield improved agreement, sensitivity, or specificity.

Our results suggest that, although variation in per-participant characteristics, such as BMI and stride length, influences the appropriate threshold to use for each individual, this variation does not seem to overtly influence the overall agreement, sensitivity, or specificity between Fitbit and ActiGraph categorizations of PA level. Additional studies are required to determine if these findings persist in a larger sample of a similar population and if these findings apply beyond our population of interest.

**AUTHOR CONTRIBUTIONS**

All authors drafted the article and revised it critically for important intellectual content, and all authors gave final approval of the version of the article to be published.

**Study conception and design.** Collins, Losina.

**Acquisition of data.** Yang, Collins, Losina.

**Analysis and interpretation of data.** Silva, Yang, Collins, Losina.

**REFERENCES**

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012;380:247–57.

2. Centers for Disease Control and Prevention. 2008 Physical Activity Guidelines for Americans: Fact sheet for Health Professionals on Physical Activity Guidelines for Older Adults. URL: https://www.cdc.gov/physicalactivity/downloads/PA_fact_sheet_olderadults.pdf.

3. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamontagne MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43:1334–59.

4. U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2nd Edition. 2018. URL: https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guide lines_2nd_edition.pdf.

5. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011;40:454–61.

6. Gorman E, Hanson HM, Yang PH, Khan KM, Liu-Ambrose T, Ashe MC. Accelerometer analysis of physical activity and sedentary behavior in older adults: a systematic review and data analysis. Eur Rev Aging Phys Act 2014;11:35–49.

7. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey 2005-2006. 2006. URL: https://www.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2005.

8. Thoma LM, Dunlop D, Song J, Lee J, Tudor-Locke C, Aguiar EJ, et al. Are older adults with symptomatic knee osteoarthritis less active than the general population? Analysis from the Osteoarthritis Initiative and NHANES. Arthritis Care Res (Hoboken) 2018;70:1448–54.

9. Lee J, Chang RW, Ehrlich-Jones L, Kwoh CK, Nevitt M, Semanik PA, et al. Sedentary behavior and physical function: objective evidence from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2015;67:666–73.

10. Rausch Osthoff AK, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77:1251–60.

11. Dunlop DD, Song J, Armstrong EK, Semanik PA, Lee J, Chang RW, et al. Sedentary time in US older adults associated with disability in activities of daily living independent of physical activity. J Phys Act Health 2015;12:93–101.

12. Dunlop DD, Song J, Lee J, Gilbert AL, Semanik PA, Ehrlich-Jones L, et al. Physical activity minimum threshold predicting improved function in adults with lower-extremity symptoms. Arthritis Care Res (Hoboken) 2017;69:475–83.

13. Tanaka R, Ozawa J, Kito N, Moriyama H. Efficacy of strengthening or aerobic exercise on pain relief in people with knee osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. Clin Rehabil 2013;27:1059–71.

14. 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.

15. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017;47:1821–45.

16. Barnett A, van den Hoek D, Barnett D, Cerin E. Measuring moderate-intensity walking in older adults using the ActiGraph accelerometer. BMC Geriatr 2016;16:211.

17. Zisko N, Nauman J, Sandbakk SB, Aspvik NP, Salvesen Ø, Carl sen T, et al. Absolute and relative accelerometer thresholds for determining the association between physical activity and metabolic syndrome in the older adults: the Generation-100 study. BMC Geriatr 2017;17:109.

18. Aguilar-Farias N, Brown WJ, Peeters GM. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. J Sci Med Sport 2014;17:293–9.

19. Santos-Lozano A, Santin-Medeiros F, Cardon G, Torres-Luque G, Bailón R, Bergmeir C, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. Int J Sports Med 2013;34:975–82.

20. Rowe DA, Welk GJ, Heil DP, Mahar MT, Kemble CD, Calabri MA, et al. Stride rate recommendations for moderate-intensity walking. Med Sci Sports Exerc 2011;43:312–8.

21. Plasqui G, Jooosen AM, Kester AD, Goris AH, Westerterp KR. Measuring free-living energy expenditure and physical activity with triaxial accelerometer. Obes Res 2005;13:1363–9.

22. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity (Silver Spring) 2007;15:2371–9.

23. Tudor-Locke C, Hatano Y, Pangrazi RP, Kang M. Revisiting “how many steps are enough?” Med Sci Sports Exerc 2008;40 Suppl:S337–43.

24. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. The appropriate threshold to use for each individual, this variation does not seem to overtly influence the overall agreement, sensitivity, or specificity. Adjusting ActiGraph MVPA thresholds by participant BMI also does not yield improved agreement, sensitivity, or specificity.

25. Murakami H, Kawakami R, Naka S, Nakata Y, Ishikawa-Takata K, Tanaka S, et al. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. JAMA Intern Med 2016;176:702–3.
26. 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.

27. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. J Med Eng Technol 2016;40:325–9.

28. Leininger LJ, Cook BJ, Jones VJ, Bellumori M, Adams KJ. Validation and accuracy of Fitbit Charge: a pilot study in a university worksite walking program: 362, Board #199, June 1, 11:00 AM - 12:30PM. Med Sci Sports Exerc 2016;48 Suppl 1:96.

29. Farina N, Lowry RG. The validity of consumer-level activity monitors in healthy older adults in free-living conditions. J Aging Phys Act 2018;26:128–35.

30. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit’s active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3:e000254.

31. 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 Exerc Med 2015;1:e000013.

32. 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.

33. Schmitt D, Vap A, Queen RM. Effect of end-stage hip, knee, and ankle osteoarthritis on walking mechanics. Gait Posture 2015;42:373–9.

34. Collins JE, Yang HY, Trentadue TP, Gong Y, Losina E. Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. PLoS One 2019;14:e0211231.

35. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.

36. Hall KS, Howe CA, Rana SR, Martin CL, Morey MC. METs and accelerometry of walking in older adults: standard versus measured energy cost. Med Sci Sports Exerc 2013;45:574–82.

37. Tudor-Locke C, Han H, Aguilar-EJ, Barreira TV, Schuna JM Jr, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. Br J Sports Med 2018;52:776–88.

38. Supakkul K. Using positional heel-marker data to more accurately calculate stride length for treadmill walking: a step length approach. arXiv: 1710.0903v1 [q-bio.QM]; 2017.

39. 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.

40. Rejeski WJ, Marsh AP, Brubaker PH, Buman M, Fielding RA, Hire D, et al. Analysis and interpretation of accelerometry data in older adults: the LIFE Study. J Gerontol A Biol Sci Med Sci 2016;71:521–8.

41. Cain KL, Conway TL, Adams MA, Husak LE, Sallis JF. Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. Int J Behav Nutr Phys Act 2013;10:51.

42. Roos EM, Toksvig-Larsen S. The Knee Injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003;1:17.