

# Accuracy of Wearable Trackers for Measuring Moderate- to Vigorous-Intensity Physical Activity: A Systematic Review and Meta-Analysis

Jessica Gorzelitz, Chloe Farber, Ronald Gangnon, and Lisa Cadmus-Bertram  
University of Wisconsin–Madison

**Background:** The evidence base regarding validity of wearable fitness trackers for assessment and/or modification of physical activity behavior is evolving. Accurate assessment of moderate- to vigorous-intensity physical activity (MVPA) is important for measuring adherence to physical activity guidelines in the United States and abroad. Therefore, this systematic review synthesizes the state of the validation literature regarding wearable trackers and MVPA. **Methods:** A systematic search of the PubMed, Scopus, SPORTDiscus, and Cochrane Library databases was conducted through October 2019 (PROSPERO registration number: CRD42018103808). Studies were eligible if they reported on the validity of MVPA and used devices from Fitbit, Apple, or Garmin released in 2012 or later or available on the market at the time of review. A meta-analysis was conducted on the correlation measures comparing wearables with the ActiGraph. **Results:** Twenty-two studies met the inclusion criteria; all used a Fitbit device; one included a Garmin model and no Apple-device studies were found. Moderate to high correlations (.7–.9) were found between MVPA from the wearable tracker versus criterion measure (ActiGraph  $n = 14$ ). Considerable heterogeneity was seen with respect to the specific definition of MVPA for the criterion device, the statistical techniques used to assess validity, and the correlations between wearable trackers and ActiGraph across studies. **Conclusions:** There is a need for standardization of validation methods and reporting outcomes in individual studies to allow for comparability across the evidence base. Despite the different methods utilized within studies, nearly all concluded that wearable trackers are valid for measuring MVPA.

**Keywords:** ActiGraph, criterion, Fitbit, fitness, validation

Accurate assessment of physical activity remains challenging. Wearable fitness trackers are ubiquitous among consumers and represent new opportunities for measurement (Kaewkannate & Kim, 2016; Lunney, Cunningham, & Eastin, 2016). Compared with research-grade devices like the ActiGraph, consumer-based wearable fitness trackers feature user-friendly applications, social features, and integrated communications across apps, which aim to promote physical activity and reduce sedentary behavior. However, these devices are most useful if they provide valid measures of activity. Within a decade, consumer devices have evolved from a simple waist-worn pedometer to integrated devices used in the home, at workplace wellness programs, in research studies, and at the individual level to help construct healthier, more active lifestyles (Block et al., 2017; Mobbs, Phan, Maharaj, & Rao, 2016; Yavelberg, Zaharieva, Cinar, Riddell, & Jamnik, 2018; Zhang, McClean, Ko, Morgan, & Schmitz, 2017).

The validity of wearable fitness trackers—including Apple Watches, Fitbits, and Garmins—has been assessed for steps, energy expenditure, moderate- to vigorous-intensity physical activity (MVPA), sedentary time, and sleep (Dooley, Golaszewski, & Bartholomew, 2017; Evenson, Goto, & Furberg, 2015; Floegel, Florez-Pregonero, Hekler, & Buman, 2017; Henriksen et al., 2018;

Huang, Xu, Yu, & Shull, 2016; Jo, Lewis, Directo, Kim, & Dolezal, 2016; Kang et al., 2017; Mantua, Gravel, & Spencer, 2016; Price et al., 2017; Roos, Taube, Beeler, & Wyss, 2017; Shcherbina et al., 2017; Toth et al., 2018; Treacy et al., 2017; Woodman, Crouter, Bassett, Fitzhugh, & Boyer, 2017). Published reviews regarding the validity of the steps feature comprise the majority of the evidence base for wearable fitness trackers (Adam Noah, Spierer, Gu, & Bronner, 2013; An, Jones, Kang, Welk, & Lee, 2017; Chandrasekar, Hensor, Mackie, Backhouse, & Harris, 2018; Treacy et al., 2017; Ummels, Beekman, Theunissen, Braun, & Beurskens, 2018). In 2015, Evenson et al. synthesized the findings of 22 published studies on the validity and reliability of wearable fitness trackers (including Fitbit and Jawbone) and the specific abilities of the devices to measure steps, distance, physical activity, energy expenditure, and sleep. At the time of Evenson's review, most studies assessed the validity of steps or energy expenditure, and only two studies assessed validity of intensities and minutes of physical activity (Evenson et al., 2015). Other systematic reviews have assessed the validity of specific applications of wearable fitness trackers, such as Coughlin and Stewart (2016), on the use of wearables in promoting physical activity (Coughlin & Stewart, 2016; O'Driscoll et al., 2018).

Publications on the validity of wearable trackers for physical activity assessment have used many different strategies, devices, and criteria for validation. To date, no systematic review has addressed minutes of activity (e.g., MVPA). This is a crucial gap because most physical activity guidelines, including those of the United States and the World Health Organization are expressed in minutes, not steps (Garriguet & Colley, 2014;

---

Gorzelitz, Farber, and Cadmus-Bertram are with the Department of Kinesiology, University of Wisconsin–Madison, Madison, WI, USA. Gangnon is with the Department of Population Health Sciences; Department of Biostatistics and Medical Informatics; and Department of Statistics, University of Wisconsin–Madison, Madison, WI, USA. Cadmus-Bertram ([lisa.bertram@wisc.edu](mailto:lisa.bertram@wisc.edu)) is corresponding author.

Physical Activity Guidelines Advisory Committee, 2018; Physical Activity Guidelines Advisory Committee, 2009). Thus, MVPA is one of the most relevant metrics for consumers and researchers. Therefore, our purpose was to systematically review validation studies published since 2012 using consumer-based wearable activity trackers to measure MVPA. We selected 2012 as the starting date to cover relevant literature since the Evenson et al. paper. Specifically, we aimed to (a) describe how MVPA is assessed against criterion measurements and (b) report the populations studied, the magnitude of agreement, and the statistical analyses used to define the validity of the wearable fitness trackers.

## Methods

### Search Strategy and Study Eligibility

A systematic review of the literature in the PubMed, Scopus, SPORTDiscus, and Cochrane Library databases was conducted in May 2018 and updated in October 2019. PROSPERO guidelines were followed (PROSPERO registration number: CRD42018103808). The search strategy was developed by identifying terminology used in previous studies (Evenson et al., 2015). The search terms were (Fitbit OR Apple [monitor OR watch OR tracker] OR Garmin) AND (validity OR validation OR validate OR comparison OR comparative OR reliability OR accuracy). The term “monitor OR watch OR tracker” needed to be added to differentiate Apple brand trackers from fruit-related literature. Searches were limited to peer-reviewed articles that were available in the English language. There was no restriction on the study population (general or targeted). One author was contacted by authors to confirm device specifics not available in the manuscript (Phillips, Petroski, & Markis, 2015). PRISMA guidelines were followed for composing and reviewing this manuscript.

### Eligible Devices

This review focused on devices manufactured by Fitbit, Garmin, or Apple, the core market leaders most often used in research studies. Other brands in the marketplace (e.g., Jawbone) were not included due to not being available for purchase at the time of the review. We did not include some of the earliest tracker models. However, the release cycle of fitness trackers is substantially faster than the pace of academic research and publishing, meaning that current models of trackers will have no peer-reviewed validation data. We therefore included device models released in 2012 or later or on the market in May 2018. The latter criterion allowed for the inclusion of a few longstanding models, such as the Fitbit Zip, that were released prior to 2012 but still on the market at the time of the literature search. Devices eligible for this review were the Fitbit Alta, Alta HR, Blaze, Charge, Charge HR, Flex, Flex 2, One, Ionic, Surge, Versa, and Zip. Garmin Vivoactive, Vivofit 2, Vivofit 3, Vivofit 4, Vivofit Jr, Vivofit Jr 2, Vivosmart, and Vivosmart HR and Apple Watch Series 1, Sport Series 1, Edition—Series 1, Series 2, Edition—Series 2, Series 3, Edition—Series 3, and Hermes. Of these eligible Fitbit, Garmin, and Apple models, validation data for MVPA minutes were only available for five Fitbit models: The Fitbit Charge HR, Fitbit Charge 2, Fitbit Zip, Fitbit Flex, and Fitbit One.

### Validation Outcomes

Studies were eligible for inclusion if they reported on the validity of at least one of the eligible devices with respect to minutes of

moderate to vigorous minutes of physical activity. Eligible studies needed to have direct comparison of MVPA from both the wearable device and the criterion measure. There were differences in the way authors chose to abstract this from the wearable fitness device, such as active (corresponding to moderate intensity) or very active (corresponding to vigorous intensity) minutes from the Fitbit, use of Fitabase database downloads, while other models/devices may have had different definitions (“Help Article: What are active minutes?” 2015; Huberty et al., 2017). Specifically, we included articles that included intensities comparable with exercise. Articles that reported validations of only steps, energy expenditure, heart rate, and so forth, were not included. The initial search strategy was not specific to the validation outcome; rather these were determined during the article selection process.

### Article Selection

After removal of duplicates, a three-phase process was used to screen the articles identified during the search. Articles that targeted special populations (e.g., children or hospitalized heart failure patients) were included throughout the selection process. The first phase of screening was based on titles only; the second phase included abstracts, and the third phase included the full-text articles. For each round, two independent reviewers assessed each article, and a third reviewer resolved any discrepancies.

### Data Extraction

The following information was abstracted from the studies: study population, study age group, proportion of men, device used, outcomes assessed, criterion used, study setting (free living or lab), type of analysis for validity, author’s definition of MVPA, and summary of findings. One reviewer extracted the data with the other two reviewers checking the information for accuracy.

### Quality Assessment

A modified version of the Downs and Black checklist was used to assess quality of the included studies (Downs & Black, 1998). The full Downs and Black checklist contains 27 items that assess three domains of reporting validity, internal, and external validity. We used a modified version consisting of 15 items for a maximum score of 15 points. Since all the studies included in this review are observational, we followed the methods used by Prince et al. and Warburton et al. using only the 15-item version of the checklist (Prince et al., 2008; Warburton, Charlesworth, Ivey, Nettlefold, & Bredin, 2010).

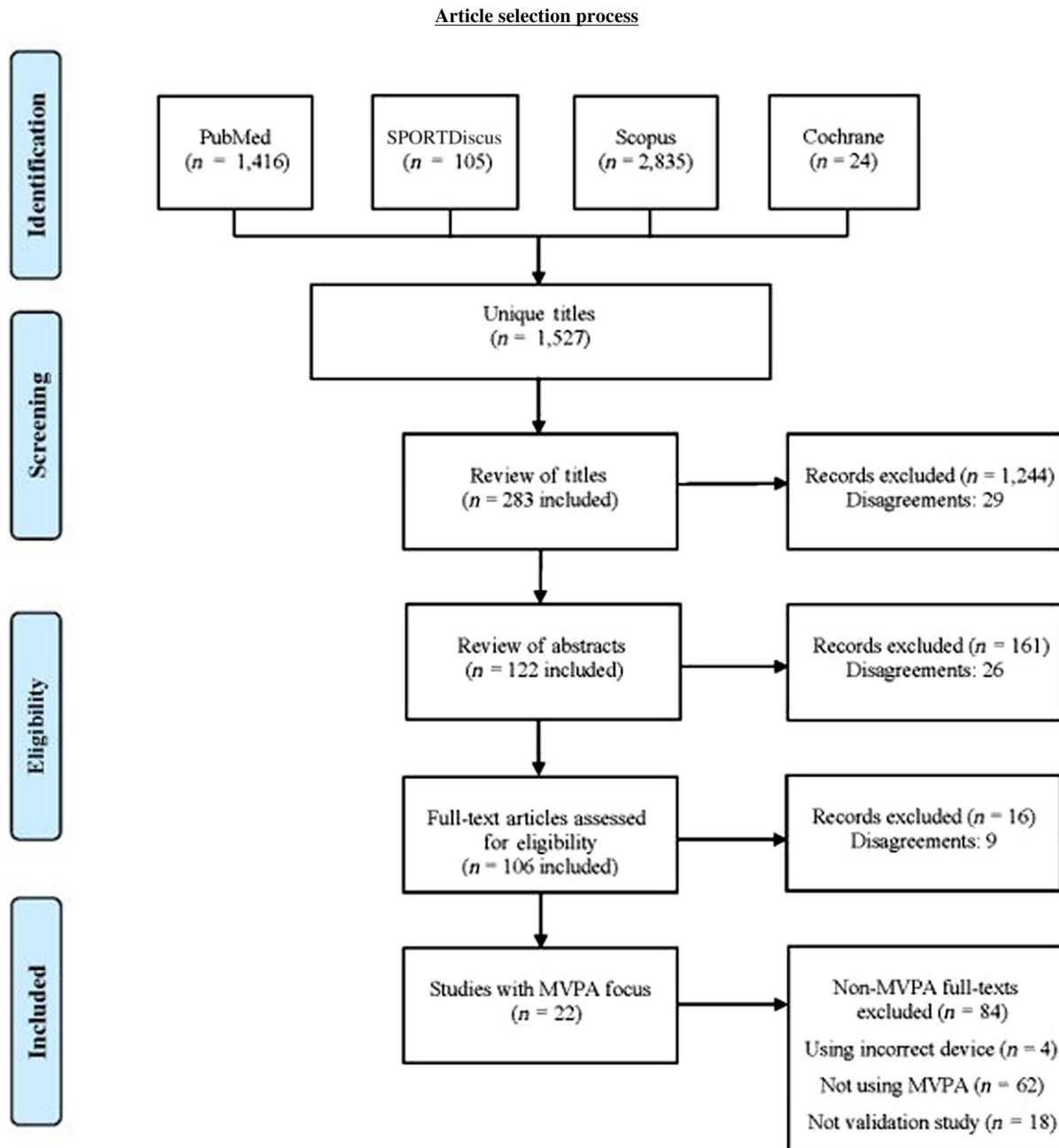
### Meta-Analysis

Correlations (Pearson’s  $r$  or Spearman’s  $\rho$ ) between the wearable fitness tracker and criterion measures of MVPA were obtained from each study. Fisher’s  $r$  to  $Z$  transformation was applied prior to analysis. Random-effects models were used to estimate the average effects and the variance of the effects across studies. The  $I^2$  statistic was used to quantify the heterogeneity of effects. A 95% confidence interval for the average correlation, and a 95% prediction interval for the correlation in a future study were calculated. Subgroup analyses based on correlation (Pearson or Spearman), age (adults or children), health (healthy adults or nonhealthy adults), location (lab based or free living), and wear location (right hip or other hip) were performed to assess potential sources of heterogeneous effects. Analyses were performed using the meta package in R studio (version 1.2.5042; R studio, Boston, MA).

## Results

The literature review identified 1,527 unique titles from the four databases (Figure 1; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). After excluding irrelevant articles, 283 papers remained. Those 283 abstracts were read by two different reviewers, a process which identified 122 articles for full-text review. Articles were excluded if they were not a validation study, if there was not a full-text manuscript (e.g., conference abstract), or if the study used a fitness tracker other than a Garmin, Fitbit, or Apple device. After full-text review, 22 unique articles focusing on the validity of MVPA were identified.

Table 1 summarizes relevant study characteristics including wearable tracker model(s), the definition of MVPA, and the type of statistical analysis. Studies included healthy ( $n = 17$ ) and disease-specific ( $n = 5$ ) populations. Studies included children and adolescents between ages 4 and 13 years ( $n = 7$ ), adults with mean ages between 28 and 43 years ( $n = 11$ ), and older populations with mean ages between 62 and 66 years ( $n = 5$ ). One study was conducted among pregnant women. Studies were typically very good quality, with a mean score of 12.6 (range 8–15) on the modified Downs and Black checklist. All studies received maximum scores for description of participant characteristics and study interventions.



**Figure 1** — Flow diagram of article selection process. MVPA = moderate- to vigorous-intensity physical activity. Adapted from “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement,” by D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, and The PRISMA Group, 2009, *PLoS Medicine*, 6(6), p. e1000097. Copyright 2009 by xxx. For more information visit [www.prisma-statement.org](http://www.prisma-statement.org).

**Table 1 Characteristics of the Validation Studies Included in this Review (n = 22<sup>a</sup>)**

<b>Author (year)</b>	<b>N</b>	<b>Population</b>	<b>% Male</b>	<b>Wearable device(s) tested</b>	<b>Criterion measure and location</b>	<b>Wear time requirement and cut points</b>	<b>How MVPA defined</b>
<b>Lab-based studies</b>							
Byun et al. (2018) <sup>a</sup>	28	Healthy children	54%	Fitbit Flex	Direct observation	N/A	Scale of 0–3 for sedentary to vigorous activity (60-s sampling): 0 ( <i>sedentary</i> ), 1 ( <i>light</i> ), 2 ( <i>moderate</i> ), and 3 ( <i>vigorous</i> )
Byun et al. (2018) <sup>a</sup>	27		59%				Time spent in “active minutes”
Imboden et al. (2017)	30	Healthy adults	50%	Fitbit Flex	AG GT3X+ (right hip)	MVPA ≥ 3 METs or 2,691+ cpm	
Kang, Kim, Byun, Suk, & Lee, (2019)	43	Healthy children	58%	Fitbit Charge HR	AG GT3X+ (both wrists)	Hand-weight exercise = 2.1 METs Moderate = 2.7–3.5 METs	Scale of 0–3 for sedentary to vigorous activity (60-s sampling): 0 ( <i>sedentary</i> ), 1 ( <i>light</i> ), 2 ( <i>moderate</i> ), and 3 ( <i>vigorous</i> )
Liang, & Getchell (2018)	35	Healthy children Healthy adults	60% 70%	Fitbit Zip	Actical	Moderate = 3–6 METs Vigorous > 6 METs	Total minutes spent in moderate and vigorous intensity
Sushames et al. (2016)	25	Healthy adults	52%	Fitbit Flex	AG GT3X+ (pre-ferred hip)	Moderate = 1,952–5,724 cpm Vigorous = 5,725–9,498 cpm Very vigorous = 9,498+ cpm	Minutes in moderate and vigorous exercise (as obtained by the Fitbit website)
<b>Free-living studies</b>							
Alharbi et al. (2016)	48	Adults w/ congestive heart disease and their families	52%	Fitbit Flex	AG GT3X (unspecified hip)	Sedentary = 0–100 cpm Light = 101–2,020 cpm Moderate = 2,021–5,999 cpm Vigorous = 6,000+ cpm MVPA = 2,021+ cpm	Fitbit: sedentary = 1 MET, moderate = 3 + METs, vigorous = 6+ METs
Brewer et al. (2017)	53	Healthy adults	17%	Fitbit Flex	AG GT3X+ (right hip)	MVPA = 1,952+ cpm	Time spent in “active minutes”
Collins, Yang, Trentadue, Gong, & Losina (2019)	15	Older adults w/ osteoarthritis	33%	Fitbit Charge2	AG GT3X+ (non-dominant wrist, unspecified hip)	Valid day ≥ 10 hr; valid week ≥ 4 days MVPA: vector magnitude ≥ 1,924 cpm in bouts of ≥ 10 min	Fitbit: MVPA > 3 METs, bouts ≥ 10 min allowing two grace minutes where steps was greater than threshold
Degroote, De Bourdeaudhuij, Verloigne, Poppe, & Crombez (2018)	36	Healthy adults	50%	Fitbit Charge	AG GT3X+ (right hip)	Valid day was 10 hr of wear Sedentary = 0–99 cpm Light = 100–1,951 cpm Moderate = 1,952–5,723 cpm Vigorous = 5,724+ cpm	Fitbit variable: “active minutes”
Dominick et al. (2016)	19	Healthy adults	21%	Fitbit Flex	AG GT3X (dominant hip)	Sedentary = 0 (0–199 cpm) Light = 1 (200–2,690 cpm) Moderate = 2 (2,691–6,166 cpm) Vigorous = 3 (6,167+ cpm)	Sedentary = 0 Light = 1 Moderate = 2 Vigorous = 3
Ferguson et al. (2015)	21	Healthy adults	48%	Fitbit One Fitbit Zip	AG GT3X+ (right hip)	Sedentary < 200 cpm Light = 200–2,690 cpm Moderate = 2,691–6,166 cpm Vigorous = 6,167+ cpm	One: sum of the very and moderate active minutes Zip: sum of the vigorous and moderate active minutes
Gomersall et al. (2016)	29	Healthy adults	52%	Fitbit One	AG GT3X+ (right hip)	Sedentary < 100 cpm MVPA = 2,020+ cpm	Time spent in “very active minutes”

(continued)

**Table 1 (continued)**

Author (year)	N	Population	% Male	Wearable device(s) tested	Criterion measure and location	Wear time requirement and cut points	How MVPA defined
Hui et al. (2018)	12	Adult stroke survivors	58%	Fitbit One	Actical (nonparetic ankle)	Metabolic: Sedentary = 1 MET Light < 3 METs Moderate = 3–6 METs Vigorous > 6 METs MVPA = 2,296+ cpm	Time spent in sedentary, light active, fairly active, and very active No specific thresholds for activity intensity were provided
Moose, Oja, Reisberg, Vilo, & Kull (2018)	147	Healthy children	50%	Fitbit Zip	AG GT3X-BT (unspecified hip)	MVPA = 2,296+ cpm	Fitbit's categories of "fairly active" was moderate and "very active" was vigorous intensity
Redenius, Kim, & Byun (2019)	65	Healthy adults	28%	Fitbit Flex	AG GT3X+ (dominant hip)	Freedson: MVPA = 1,952+ cpm Troiano: MVPA = 2,020+ cpm VM3: MVPA = 2,691+ cpm	Downloaded from Fitabase, minute by minute; used the Fitbit algorithm of 2 = moderate PA, 3 = vigorous PA
Rosenberger et al. (2016)	40	Healthy adults	48%	Fitbit One	AG GT3X+ (right hip)	MVPA = 1,580+ cpm	Minutes in moderate plus vigorous physical activity
Schneider and Chau (2016)	25 35 27	Healthy adolescents (three cohorts)	48% 47% 40%	Fitbit Zip	AG GT3X (unspecified hip)	MVPA ≥ 4 METs Cohort 1 + 3 MVPA: 2,058 cpm Cohort 2: 2,220 cpm	Total time in "active" and "very active"
St-Laurent, Mony, Mathieu, & Ruchat (2018)	16	Pregnant women	0%	Fitbit Zip Fitbit Flex	AG GT3X (unspecified hip)	Valid wear of 10 hr for 4 days Moderate = 1,952–5,724 cpm Vigorous = 5,725 cpm	Not reported
Sushames et al. (2016)	25	Healthy adults	52%	Fitbit Flex	AG GT3X+ (preferred hip)	Moderate = 1,952–5,724 cpm Vigorous = 5,725–9,498 cpm Very vigorous = 9,498+ cpm	Minutes in moderate and vigorous exercise (as obtained by the Fitbit website)
Tedesco et al. (2019)	20	Older, healthy adults	45%	Fitbit Charge2, Garmin vivosmartHR+	AG GT9X-BT (dominant hip)	Valid day ≥ 10 hr MVPA = 2,020+ cpm	Not reported
Van Blarigan et al. (2017)	25	Adults w/ prostate cancer	100%	Fitbit One	AGGT3X+ (right hip)	Light = 100–2,019 cpm Moderate = 5,725–9,498 cpm Vigorous = 9,498+ cpm	Light < 3 METs, moderate 3–5.9 METs, and vigorous ≥ 6 METs
Voss et al. (2017)	30	Children w/ congenital heart disease	47%	Fitbit Charge HR	AG GT3X+ (right hip)	MVPA = 2,296+ cpm	Fairly + very active minutes

Note. AG = ActiGraph; MET = metabolic equivalent of task; cpm = counts per minute; HR = heart rate; MVPA = moderate- to vigorous-intensity physical activity; N/A = not applicable; PA = physical activity; VM3 = composite vector magnitude (VM) of activity counts from all three (3) axes from the accelerometer.

\*Two separate publications in different journals, but near identical results, methods, and study.

Criterion assessments included multiple ActiGraph models, direct observation, Actical accelerometer, or indirect calorimetry (converting energy expenditure to physical activity intensity for comparison with wearable fitness device data). Of the 20 studies using the ActiGraph, 13 used the GT3X+, five used the GT3X, two used the Bluetooth-enabled GT3X-BT, one used the GT9X, and one used the GT9X-BT. The ActiGraph was not always worn on the right side on a belt. Of the 15 studies (68%) that had a sample size fewer than 40, four (18%) had fewer than 20 subjects.

A variety of statistical techniques were used to assess the validity between wearable fitness trackers and the criterion measure (Table 2). The most frequently used statistical assessments were Bland–Altman plots ( $n = 15$ ) and Pearson or Spearman correlation ( $n = 14$ ). Other approaches were mean absolute percent error ( $n = 8$ ), regression ( $n = 8$ ), and  $t$  tests for mean differences ( $n = 10$ ). Papers typically used multiple statistical approaches with an average of 4.8 statistical tests per manuscript and a maximum of eight.

For studies that utilized correlation, Table 3 shows the type of correlation (Spearman, Pearson, and intraclass correlation coefficient) and the strength of association reported. In studies focused on healthy children, the correlation values for the trackers against the criterion measure ranged between .7 (Schneider & Chau, 2016) and .9 when using direct observation as the criterion (Byun, Lee, Kim, & Brusseau, 2018). One study examined validity in a clinical-based sample of children; Voss et al. found an intraclass correlation coefficient of less than .7 in a sample of children with congenital heart disease (Voss, Gardner, Dean, & Harris, 2017). Overall, authors reported high correlations between the Fitbit Zip and the ActiGraph and concluded that the Fitbit Flex was accurate for sedentary behavior and for total physical activity in children.

Like children's populations, validation studies among adults were conducted in both clinical and healthy populations with generally moderate to high (.5–.9) correlations observed. In a study of adults with congestive heart failure and their families, authors found correlations of .7 between the Fitbit Flex and the ActiGraph, with lower correlations for women (.6) than men (.8) (Alharbi, Bauman, Neubeck, & Gallagher, 2016). In studies using the Fitbit Flex against the ActiGraph as the criterion measure, "active minutes" was correlated with the criterion measure measured at .7 (Brewer, Swanson, & Ortiz, 2017). Other studies including the Fitbit Flex found underestimations of both active minutes (Imboden, Nelson, Kaminsky, & Montoye, 2017) and time in sedentary and light activity (Dominick, Winfree, Pohlig, & Papas, 2016), while other studies found that the Fitbit was more accurate for specifically studied moderate and vigorous intensity activities of jogging and stair stepping while underestimating steps (Sushames, Edwards, Thompson, McDermott, & Gebel, 2016). For those studies that used the Fitbit One or the Fitbit Zip against the ActiGraph, MVPA correlations ranged from .7 to .9 (Ferguson, Rowlands, Olds, & Maher, 2015; Gomersall et al., 2016) while another study reported high error rates with MVPA measurement between 52% and 92% (Rosenberger, Buman, Haskell, McConnell, & Carstensen, 2016).

The results of the meta-analysis showed a mean correlation of .74 (.66, .81) across all studies assessing MVPA and using correlation analyses. High heterogeneity was noted across studies with an  $I^2$  of 43% ( $p = .05$ ). To further explore the heterogeneity across studies, several moderators were investigated as described in the methods. None of these moderators were statistically significant predictors showing meaningful heterogeneity across the selected strata. The 95% prediction interval for correlation analyses was found to be .39–.91, which is shown in Figure 2.

## Discussion

The most used device manufacturer was the Fitbit, with ActiGraph as the most common criterion device. Across different populations using different wearable devices, moderate to high correlations were found between MVPA on the wearable and the criterion assessment. While Fitbit is a propriety company, some researchers relied on "active" minutes or "very active" minutes, where others used the categorical definitions provided from a Fitabase download. While the correlations were mostly moderate to high across studies, the definition of "sufficient" correlation depends heavily on the context and the required level of precision for a specific use. Consumers will have different expectations for what is useful compared with a researcher hoping to reply on a precise, physiological measurement.

Each study differed with respect to the metric used to report agreement between the Fitbit and the criterion measure. These analytic/methodological differences made it very hard to compare across studies. As seen in Table 1, there were several different ActiLife cut points applied to adult populations leading to different definitions of MVPA based on counts per minute. Authors differed with respect to what they considered "good" validity. For example, two studies that found a correlation of .8 have reported both "excellent" and "adequate" agreement between devices (Alharbi et al., 2016; Van Blarigan et al., 2017). Furthermore, correlation analyses were the most frequently utilized statistical technique, which is not the most appropriate technique for assessing validity. Correlation analyses are highly influenced by the sample selection process. A recent publication by Welk et al. has recommended standardized protocols for future validation studies including valid, appropriate statistical techniques (Welk et al., 2019).

Statistical validity was assessed using various techniques to either describe the associations between the devices (such as correlation analyses), to describe the differences between the devices, such as mean or relative percentage error, or choosing to report the bias between devices, such as using a Bland–Altman plot. The many different statistical techniques are highlighted in Table 2. Only one article (Rosenberger et al., 2016) used equivalence testing, which has the potential to establish the unbiased accuracy of the measurement. While equivalence testing has the capacity to examine the unbiasedness between measurements, the method is infrequently used with most authors using correlations. Even within correlation analyses alone, Table 3 shows the many ways authors chose to compare MVPA with other measures. While Pearson correlations were more common, many studies have sample sizes of fewer than 40 ( $n = 15$ ) and the required tests for normality were not always explicitly stated in the methods section of those publications. These small sample sizes are especially important to consider regarding the statistical analyses, where the normal distribution assumption is probably unsatisfied and thus, a Pearson correlation or Bland–Altman analyses requiring a 95% confidence interval constructed on a normal distribution will influence the results.

Studies that reported a systematic bias appeared to find that wearables tended to overestimate (vs. underestimate) MVPA relative to the ActiGraph. Possible reasons include that the Fitbit overestimates activity based on algorithmic decisions made within the device/the data processing. However, it is possible that wrist-worn accelerometry overestimates activity and that perhaps the issue is the comparison between a wrist-worn consumer's wearable versus a hip-worn research-grade device but this issue should persist across all wrist-worn devices. However, as Table 1 shows, not every validation study followed the same protocols for ActiGraph wear with some devices worn on the wrist, some on the dominant hip, some on the

**Table 2 List of Statistical Analyses Used Within and Across Studies**

Study	Correlation (r/rho)	ICC	Regression	Kappa	Mean absolute percentage error	Mean relative error	Mean absolute difference	Sensitivity	Specificity	PPV	Bland-Altman	Coefficient of variation	ROC-AUC	t tests	RM ANOVA	Equivalence testing
Alharbi et al. (2016)	✓		✓					✓	✓	✓	✓		✓	✓		
Brewer et al. (2017)	✓		✓								✓			✓		
Byun et al. (2018)	✓			✓	✓			✓					✓			
Byun et al. (2018)	✓			✓	✓			✓					✓			✓
Collins et al. (2019)		✓														
Degroote et al. (2018)	✓		✓								✓					
Dominick et al. (2016)	✓			✓										✓		
Ferguson et al. (2015)	✓						✓									
Gomersall et al. (2016)	✓		✓		✓						✓					
Hui et al. (2018)	✓			✓												
Imboden et al. (2017)	✓			✓		✓					✓			✓		✓
Kang et al. (2019)	✓			✓	✓			✓			✓		✓	✓		✓
Liang, & Getchell (2018)														✓ <sup>a</sup>		✓
Mooses et al. (2018)	✓										✓					
Redenius et al. (2019)	✓			✓	✓						✓			✓		✓ <sup>b</sup>
Rosenberger et al. (2016)				✓							✓					
Schneider and Chau (2016)	✓										✓			✓		
St-Laurent et al. (2018)											✓			✓		✓
Sushames et al. (2016)		✓			✓						✓		✓	✓		
Tedesco et al. (2019)		✓			✓		✓				✓					
Van Blarigan et al. (2017)	✓										✓					
Voss et al. (2017)								✓			✓		✓			

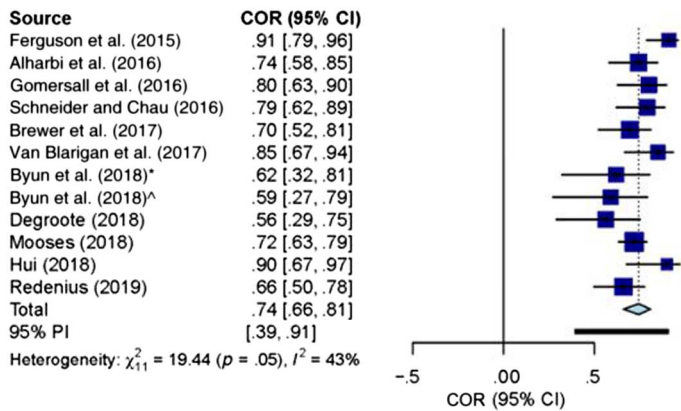
Note. ICC = intraclass correlation coefficient; ANOVA = analysis of variance; RM = repeated measures; AOC-ROC = area under the curve—receiver operator curve; PPV = positive predictive value. <sup>a</sup>Chi-square goodness of fit. <sup>b</sup>One-way ANOVA.

**Table 3 Magnitude of Association Observed in Studies that Reported Correlations**

Author (year)	Metric	Correlation magnitude
Studies comparing Fitbit with ActiGraph		
Alharbi et al. (2016)	Moderate Vigorous MVPA	Moderate: $r = .76$ Vigorous: $r = .19$ MVPA: $r = .74$ (also available in paper: results stratified by sex and for cardiac patients only)
Brewer et al. (2017)	MVPA and “active minutes”	MVPA: $r = .70$ “active minutes”: $r = .66$
Byun et al. (2018)	MVPA and total PA (two algorithms used)	Pate: MVPA: $r = .59$ , total PA: $r = .56$ Evenson: MVPA: $r = .58$ , total PA: $r = .49$
Degroote et al. (2018)	MVPA	$\rho = .56$ , ICC = <b>.66</b>
Dominick et al. (2016)	Moderate and vigorous	Moderate: $r = .43$ Vigorous: $r = .80$
Ferguson et al. (2015)	MVPA	Fitbit Zip: $r = .88$ , ICC = .36 Fitbit One: $r = .91$ , ICC = .46
Gomersall et al. (2016)	MVPA	$r/\rho = .80$ , ICC = <b>.72</b>
Imboden et al. (2017)	Fitbit’s “active minutes”	$r = .10$
Kang et al. (2019)	MVPA	Dominant vs. nondominant wrist ICC = <b>.77</b> Crouter: ICC = <b>.73</b> Chandler: ICC = .28
Mooses et al.(2018)	MVPA	Class time: $\rho = .24$ PE lesson: $\rho = .72$ Recess: $\rho = .56$
Redenius et al. (2019)	MVPA	Freedson: $r = .66$ , $\rho = .71$ Troiano: $r = .65$ , $\rho = .69$ VM3: $r = .76$ , $\rho = .79$
Schneider and Chau (2016)	MVPA	Cohort 1: $r = .67$ Cohort 2: $r = .79$ Cohort 3: $r = .94$
Sushames et al. (2016)	Specific modalities of PA	Walking on incline: $-0.06, 0.21, 0.02$ Jogging: <b>0.58, 0.38, 0.52</b> Stair stepping: <b>0.69, 0.43, 0.72</b>
Van Blarigan et al. (2017)	MVPA, moderate, and vigorous	MVPA: $r = .85$ Vigorous: $\rho = .65$ Moderate $r = .70$
Voss et al. (2017)	MVPA	Full sample: $r = .54$ , ICC = .66 Boys: $r = .72$ , ICC = .82 Girls: $r = .38$ , ICC = .49 Age < 13 years: $r = .41$ , ICC = .52 Age > 13 years: $r = .70$ , ICC = .78
Studies comparing Fitbit with Actical		
Hui et al. (2018)	Moderate and vigorous	Moderate: $r = .90, .91, .83$ Vigorous: $r = .86$
Studies comparing Fitbit with Garmin		
Tedesco et al. (2019)	MVPA	ICC = <b>.96</b>
Studies not using correlations		
Collins et al. (2019)	Pairwise comparisons of MVPA	Fitbit underestimated MVPA by 5 min
Liang, & Getchell (2018)	Chi-square goodness of fit	Fitbit indicated more time in moderate/vigorous activity
Rosenberger et al. (2016)	Bland–Altman	No systematic bias found in MVPA
St-Laurent et al. (2018)	Bland–Altman	No systematic bias found in MVPA

Note. Bold indicates statistical significance. Correlations were not reported by the following studies: Collins et al. (2019), Liang, & Getchell (2018), Rosenberger et al. (2016), and St-Laurent et al. (2018). ICC = intraclass correlation coefficient; MVPA = moderate- to vigorous-intensity physical activity; PA = physical activity; PE = physical education.





**Figure 2** — Forest plot of included studies used for meta-analysis ( $n = 12$ ) (Byun, Kim, Brusseau, 2018). COR = correlation; CI = confidence interval; PI = prediction interval.

right hip and others unspecified. Finally, it is possible that participants in free-living studies might have better adherence to the easy to wear commercial wearables versus the more cumbersome Acti-Graph. While this is unlikely, it is a possible explanation for the overreporting found in many of the studies.

While this review provides context and description of the current state of the validation evidence for wearable fitness trackers, there are still several areas for future investigation and improvement. Beyond MVPA, wearable fitness trackers can also measure light-intensity activity as well as sedentary behaviors, which contribute to energy balance and health outcomes (Rosenberger et al., 2019). While a few studies in our review have included measures of light physical activity or sedentary behavior, a preliminary review of the literature shows only three systematic reviews on this topic including one review on the health benefits associated with light-intensity physical activity (Fuezeki, Engeroff, & Banzer, 2017). Two reviews that focus on sedentary behavior examine the validity of device-based measures and the role of motion sensing technology in relationship to physical activity and sedentary behavior (Gierisch et al., 2015; Heesch, Hill, Aguilar-Farias, van Uffelen, & Pavey, 2018). While this is not a systematic review of the literature, future work should examine the independent and combined contributions of wearable fitness trackers beyond MVPA including light-intensity physical activity and sedentary behavior.

We used the modified Downs and Black checklist to assess the methodological quality of the included studies. The included studies were generally of high quality, with a mean score of 12.6 out of 15 items for validity. The risk of bias in individual studies with respect to recruitment strategies is difficult to assess especially for individual validation studies. One potential source of bias is the use of different cut points for MVPA or the highly variable definitions across devices when it comes to intensities of activity. These across-study differences are captured in Table 1.

The meta-analysis provided an average correlation across the included studies, but also showed a high level of variation across studies. This variation is to be expected with the high variability of the methodological choices represented in Table 1, as well as the use of correlation analyses. As previously mentioned, correlation is not ideal for assessing validity especially when the analysis is very sensitive to both the population sampling strategy and the inherent variation of MVPA within the studied population. We chose five meaningful moderators to explain some of the underlying

variability, but those moderators did not explain the heterogeneity across studies. Previous studies have shown imperfect agreement between different methods of activity assessment, such as Prince et al. showing a mean agreement of 0.8 between self-reported and device-based measurements (Prince et al., 2008). However, our meta-analysis shows that there is a consistent, positive, moderate (0.74) relationship between the criterion measure and the wearable fitness tracker. Determining if a correlation of .74, on average, is sufficient for measuring MVPA certainly depends on the context. At a population level, this may be appropriate; however, if these devices are used for an individual's measurement of performance that level of association may be inadequate. The heterogeneity observed in these included studies is not easily explained; hence future research is warranted to better understand the relationship between these two measures, ideally using standardized validation protocols to enhance comparability.

Future work needs to include population subgroups, such as the very young, the very old, those with mobility limitations or gait difference, and those with other behavioral patterns related to chronic health or disability. This focus will become increasingly important as our population ages and more wearables are on market targeted toward children. Additionally, there is a need for standardization in reporting of outcomes (Montoye, Moore, Bowles, Korycinski, & Pfeiffer, 2018; Welk, 2019; Welk et al., 2019). As previously mentioned, there was high dissimilarity across studies, which makes aggregating results challenging. Future work should consider the use of multiple approaches to increase the comparability of results across papers, such as included the mean absolute percentage error and a Bland-Altman plot for example. Finally, future work should focus on non-Fitbit trackers as well. The state of the evidence has well-represented Fitbit, and other manufacturers need to be included to provide a complete picture for researchers and consumers alike.

This systematic review has both strengths and limitations. Our review focuses on MVPA, which is the intensity of activity that forms the basis for the national and international physical activity guidelines—but which is an area that has not been addressed by previous reviews. Additionally, our search criteria examined a wide distribution of study types, examining populations across the lifespan as well as healthy and chronic disease populations. However, there were a relatively small number of papers found despite the wide net. Consequently, summarizing the results across studies was challenging due to the wide variety of analytic methods used, criterion definitions, and locations of the criterion. We also were able to conduct a meta-analysis to examine the mean correlation across included studies, even though the correlation analyses are not the strongest assessment of validity. Finally, due to the pace of technology versus science, several papers reported on tracker models that are no longer on the market. This technology pace is a consideration to make for future research as well as consumers evaluating the appropriate wearable device. Nevertheless, this review reports on the state of the evidence regarding validation studies in wearable fitness trackers and assessing MVPA, showing the relatively high (.74) correlations between the criterion measures and these consumer-based devices. Standardization of validation methods and reporting outcomes in individual studies is necessary to allow for comparability across the evidence base.

## References

- Adam Noah, J., Spierer, D.K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *Journal of Medical*

- Engineering & Technology*, 37(7), 456–462. PubMed ID: 24007317 doi:10.3109/03091902.2013.831135
- Alharbi, M., Bauman, A., Neubeck, L., & Gallagher, R. (2016). Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. *European Journal of Preventive Cardiology*, 23(14), 1476–1485. PubMed ID: 26907794 doi:10.1177/2047487316634883
- An, H.-S., Jones, G.C., Kang, S.-K., Welk, G.J., & Lee, J.-M. (2017). How valid are wearable physical activity trackers for measuring steps? *European Journal of Sport Science*, 17(3), 360–368. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=121039493&site=ehost-live&scope=site>
- Block, V.J., Lizee, A., Crabtree-Hartman, E., Bevan, C.J., Graves, J.S., Bove, R., . . . Gelfand, J.M. (2017). Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *Journal of Neurology*, 264(2), 316–326. PubMed ID: 27896433 doi:10.1007/s00415-016-8334-6
- Brewer, W., Swanson, B.T., & Ortiz, A. (2017). Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. *BMJ Open Sport & Exercise Medicine*, 3(1), e000254. PubMed ID: 29018543 doi:10.1136/bmjsem-2017-000254
- Byun, W., Kim, Y., & Brusseau, T.A. (2017). The Use of a Fitbit Device for Assessing Physical Activity and Sedentary Behavior in Preschoolers. *The Journal of Pediatrics*, 199, 35–40. PubMed ID: 29754862 doi:10.1016/j.jpeds.2018.03.057
- Byun, W., Lee, J.M., Kim, Y., & Brusseau, T.A. (2018). Classification accuracy of a wearable activity tracker for assessing sedentary behavior and physical activity in 3–5-year-old children. *International Journal of Environmental Research and Public Health*, 15(4). doi:10.3390/ijerph15040594
- Chandrasekar, A., Hensor, E.M.A., Mackie, S.L., Backhouse, M.R., & Harris, E. (2018). Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica. *Gait & Posture*, 61, 339–345. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=128611076&site=ehost-live&scope=site>
- Coughlin, S.S., & Stewart, J. (2016). Use of consumer wearable devices to promote physical activity: A review of health intervention studies. *Journal of Environment and Health Sciences*, 2(6). doi:10.3390/ijerph15040594
- Collins, J.E., Yang, H.Y., Trentadue, T.P., Gong, Y., & Losina, E. (2019). Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. *PLoS One*, 14(1), e0211231. PubMed ID: 30477509 doi:10.1371/journal.pone.0211231
- Degroote, L., De Bourdeaudhuij, I., Verloigne, M., Poppe, L., & Crombez, G. (2018). The Accuracy of Smart Devices for Measuring Physical Activity in Daily Life: Validation Study. *JMIR Mhealth Uhealth*, 6(12), e10972. PubMed ID: 30545810 doi:10.2196/10972
- Dominick, G.M., Winfree, K.N., Pohlrig, R.T., & Papas, M.A. (2016). Physical activity assessment between consumer- and research-grade accelerometers: A comparative study in free-living conditions. *JMIR mHealth uHealth*, 4(3), e110. PubMed ID: 27644334 doi:10.2196/mhealth.6281
- Dooley, E.E., Golaszewski, N.M., & Bartholomew, J.B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR mHealth uHealth*, 5(3), e34. PubMed ID: 28302596 doi:10.2196/mhealth.7043
- Downs, S.H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, 52(6), 377–384. PubMed ID: 9764259 doi:10.1136/jech.52.6.377
- Evenson, K.R., Goto, M.M., & Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 159. PubMed ID: 26684758 doi:10.1186/s12966-015-0314-1
- Ferguson, T., Rowlands, A.V., Olds, T., & Maher, C. (2015). 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*, 12(1), 42. PubMed ID: 25890168 doi:10.1186/s12966-015-0201-9
- Floegel, T.A., Florez-Pregonero, A., Hekler, E.B., & Buman, M.P. (2017). Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 72(2), 229–236. PubMed ID: 27257217 doi:10.1093/gerona/glw098
- Fuezeki, E., Engeroff, T., & Banzer, W. (2017). Health benefits of light-intensity physical activity: A systematic review of accelerometer data of the National Health and Nutrition Examination Survey (NHANES). *Sports Medicine*, 47(9), 1769–1793.
- Garriguet, D., & Colley, R.C. (2014). A comparison of self-reported leisure-time physical activity and measured moderate-to-vigorous physical activity in adolescents and adults. *Public Health Reports*, 25(7), 3–11.
- Gierisch, J.M., Goode, A.P., Batch, B.C., Huffman, K.N., Hall, K.S., Hastings, S.N., . . . Williams, J.W., Jr. (2015). VA evidence-based synthesis program reports. In *The impact of wearable motion sensing technologies on physical activity: A systematic review*. Washington, DC: U.S. Department of Veterans Affairs.
- Gomersall, S.R., Ng, N., Burton, N.W., Pavey, T.G., Gilson, N.D., & Brown, W.J. (2016). Estimating physical activity and sedentary behavior in a free-living context: A pragmatic comparison of consumer-based activity trackers and actigraph accelerometry. *Journal of Medical Internet Research*, 18(9), e239. PubMed ID: 27604226 doi:10.2196/jmir.5531
- Heesch, K.C., Hill, R.L., Aguilar-Farias, N., van Uffelen, J.G.Z., & Pavey, T. (2018). Validity of objective methods for measuring sedentary behaviour in older adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 15(1), 119. PubMed ID: 30477509 doi:10.1186/s12966-018-0749-2
- Help Article: What are active minutes. (2015). *Fitbit*. Retrieved from [http://help.fitbit.com/articles/en\\_US/Help\\_article/What-are-very-active-minutes/](http://help.fitbit.com/articles/en_US/Help_article/What-are-very-active-minutes/)
- Henriksen, A., Haugen Mikalsen, M., Woldaregay, A.Z., Muzny, M., Hartvigsen, G., Hopstock, L.A., & Grimsgaard, S. (2018). Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research*, 20(3), e110. PubMed ID: 29567635 doi:10.2196/jmir.9157
- Huang, Y., Xu, J., Yu, B., & Shull, P.B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait & Posture*, 48, 36–41. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=116906805&site=ehost-live&scope=site>
- Huberty, J.L., Buman, M.P., Leiferman, J.A., Bushar, J., Hekler, E.B., & Adams, M.A. (2017). Dose and timing of text messages for increasing physical activity among pregnant women: A randomized controlled trial. *Translational Behavioral Medicine*, 7(2), 212–223. PubMed ID: 27800565 doi:10.1007/s13142-016-0445-1
- Hui, J., Heyden, R., Bao, T., Accettone, N., McBay, C., Richardson, J., & Tang, A. (2018). Validity of the Fitbit One for Measuring Activity in Community-Dwelling Stroke Survivors. *Physiotherapy Canada*, 70(1), 81–89. PubMed ID: 29434422 doi:10.3138/ptc.2016-40.ep

- Imboden, M.T., Nelson, M.B., Kaminsky, L.A., & Montoye, A.H. (2017). Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. *British Journal of Sports Medicine*. doi:10.1136/bjsports-2016-096990
- Jo, E., Lewis, K., Directo, D., Kim, M.J., & Dolezal, B.A. (2016). Validation of biofeedback wearables for photoplethysmographic heart rate tracking. *Journal of Sports Science and Medicine*, 15(3), 540–547. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=117408634&site=ehost-live&scope=site>
- Kaewkannate, K., & Kim, S. (2016). A comparison of wearable fitness devices. *BMC Public Health*, 16(1), 433. PubMed ID: 27220855 doi:10.1186/s12889-016-3059-0
- Kang, S.G., Kang, J.M., Ko, K.P., Park, S.C., Mariani, S., & Weng, J. (2017). Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *Journal of Psychosomatic Research*, 97, 38–44. PubMed ID: 28606497 doi:10.1016/j.jpsychores.2017.03.009
- Kang, S., Kim, Y., Byun, W., Suk, J., & Lee, J.M. (2019). Comparison of a Wearable Tracker with Actigraph for Classifying Physical Activity Intensity and Heart Rate in Children. *International Journal of Environmental Research and Public Health*, 16(15), 2663. PubMed ID: 31349667 doi:10.3390/ijerph16152663
- Liang, L.Y., & Getchell, N. (2018). A preliminary comparison of physical activity intensity classification and step counts between the Fitbit Zip and Actical accelerometer during treadmill locomotion. *Gazzetta Medica Italiana Archivio per le Scienze Mediche*, 177(9), 417–424.
- Lunney, A., Cunningham, N.R., & Eastin, M.S. (2016). Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior*, 65, 114–120. doi:10.1016/j.chb.2016.08.007
- Mantua, J., Gravel, N., & Spencer, R.M. (2016). Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors (Basel)*, 16(5): 646. doi:10.3390/s16050646
- Mobbs, R.J., Phan, K., Maharaj, M., & Rao, P.J. (2016). Physical activity measured with accelerometer and self-rated disability in lumbar spine surgery: A prospective study. *Global Spine Journal*, 6(5), 459–464. PubMed ID: 27433430 doi:10.1055/s-0035-1565259
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(6), e1000097. doi:10.1371/journal.pmed.1000097
- Montoye, A.H., Moore, R.W., Bowles, H.R., Korycinski, R., & Pfeiffer, K.A. (2018). Reporting accelerometer methods in physical activity intervention studies: A systematic review and recommendations for authors. *British Journal of Sports Medicine*, 52(23), 1507–1516. PubMed ID: 27539504 doi:10.1136/bjsports-2015-095947
- Mooses, K., Oja, M., Reisberg, S., Vilo, J., & Kull, M. (2018). Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study. *BMC Public Health*, 18(1), 858. PubMed ID: 29996797 doi:10.1186/s12889-018-5752-7
- O'Driscoll, R., Turicchi, J., Beaulieu, K., Scott, S., Matu, J., Deighton, K., ... Stubbs, J. (2018). How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine*. doi:10.1136/bjsports-2018-099643
- Phillips, L.J., Petroski, G.F., & Markis, N.E. (2015). A comparison of accelerometer accuracy in older adults. *Research in Gerontological Nursing*, 8(5), 213–219. PubMed ID: 25942386 doi:10.3928/19404921-20150429-03
- Physical Activity Guidelines Advisory Committee. (2018). *Physical activity guidelines advisory committee scientific report (F2–33)*. Washington, DC: US Department of Health and Human Services.
- Physical Activity Guidelines Advisory Committee Report 2008. To the Secretary of Health and Human Services. Part A: Executive summary. (2009). *Nutrition Reviews*, 67(2), 114–120. doi:10.1111/j.1753-4887.2008.00136.x
- Price, K., Bird, S.R., Lythgo, N., Raj, I.S., Wong, J.Y., & Lynch, C. (2017). Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering & Technology*, 41(3), 208–215. PubMed ID: 27919170 doi:10.1080/03091902.2016.1253795
- Prince, S.A., Adamo, K.B., Hamel, M.E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 5(1), 56. PubMed ID: 18990237 doi:10.1186/1479-5868-5-56
- Redenius, N., Kim, Y., & Byun, W. (2019). Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC Medical Research Methodology*, 19(1), 29. PubMed ID: 30732582 doi:10.1186/s12874-019-0668-1
- Roos, L., Taube, W., Beeler, N., & Wyss, T. (2017). Validity of sports watches when estimating energy expenditure during running. *BMC Sports Science, Medicine & Rehabilitation*, 9, 1–8. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=127038926&site=ehost-live&scope=site>
- Rosenberger, M.E., Buman, M.P., Haskell, W.L., McConnell, M.V., & Carstensen, L.L. (2016). Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine & Science in Sports & Exercise*, 48(3), 457–465. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=sph&AN=113219482&site=ehost-live&scope=site>
- Rosenberger, M.E., Fulton, J.E., Buman, M.P., Troiano, R.P., Grandner, M.A., Buchner, D.M., & Haskell, W.L. (2019). The 24-hour activity cycle: A new paradigm for physical activity. *Medicine & Science in Sports & Exercise*, 51(3), 454–464. PubMed ID: 30339658 doi:10.1249/MSS.0000000000001811
- Schneider, M., & Chau, L. (2016). Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Research Notes*, 9(1), 448. PubMed ID: 27655477 doi:10.1186/s13104-016-2253-6
- Shcherbina, A., Mattsson, C.M., Waggott, D., Salisbury, H., Christle, J.W., Hastie, T., ... Ashley, E.A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, 7(2), 3. doi:10.3390/jpm7020003
- St-Laurent, A., Mony, M.M., Mathieu, M.E., & Ruchat, S.M. (2018). Validation of the Fitbit Zip and Fitbit Flex with pregnant women in free-living conditions. *Journal of Medical Engineering & Technology*, 42(4), 259–264. PubMed ID: 30198806 doi:10.1080/03091902.2018.1472822
- Sushames, A., Edwards, A., Thompson, F., McDermott, R., & Gebel, K. (2016). Validity and reliability of Fitbit flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One*, 11(9), e0161224. PubMed ID: 27589592 doi:10.1371/journal.pone.0161224
- Tedesco, S., Sica, M., Ancillao, A., Timmons, S., Barton, J., & O'Flynn, B. (2019). Validity Evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in Free-Living Environments in an Older Adult Cohort. *Journal of Medical Internet Research Mhealth Uhealth*, 7(6), e13084. PubMed ID: 31219048 doi:10.2196/13084

- Toth, L.P., Park, S., Springer, C.M., Feyerabend, M.D., Steeves, J.A., & Bassett, D.R. (2018). Video-recorded validation of wearable step counters under free-living conditions. *Medicine & Science in Sports & Exercise*, *50*(6), 1315–1322. PubMed ID: 29381649 doi:10.1249/MSS.0000000000001569
- Treacy, D., Hassett, L., Schurr, K., Chagpar, S., Paul, S.S., & Sherrington, C. (2017). Validity of different activity monitors to count steps in an inpatient rehabilitation setting. *Physical Therapy*, *97*(5), 581–588. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip.uid&db=sph&AN=123661185&site=ehost-live&scope=site>
- Ummels, D., Beekman, E., Theunissen, K., Braun, S., & Beurskens, A.J. (2018). Counting steps in activities of daily living in people with a chronic disease using nine commercially available fitness trackers: Cross-sectional validity study. *JMIR mHealth uHealth*, *6*(4), e70. PubMed ID: 29610110 doi:10.2196/mhealth.8524
- Van Blarigan, E.L., Kenfield, S.A., Tatum, L., Cadmus-Bertram, L.A., Carroll, P.R., & Chan, J.M. (2017). The Fitbit One physical activity tracker in men with prostate cancer: Validation study. *JMIR Cancer*, *3*(1), e5. doi:10.2196/cancer.6935
- Voss, C., Gardner, R.F., Dean, P.H., & Harris, K.C. (2017). Validity of commercial activity trackers in children with congenital heart disease. *Canadian Journal of Cardiology*, *33*(6), 799–805. PubMed ID: 28347581 doi:10.1016/j.cjca.2016.11.024
- Warburton, D.E., Charlesworth, S., Ivey, A., Nettlefold, L., & Bredin, S.S. (2010). A systematic review of the evidence for Canada's physical activity guidelines for adults. *International Journal of Behavioral Nutrition and Physical Activity*, *7*(1), 39. doi:10.1186/1479-5868-7-39
- Welk, G.J. (2019). Harmonizing monitor-and report-based estimates of physical activity through calibration. *Kinesiology Review*, *8*(1), 16–24. doi:10.1123/kr.2018-0064
- Welk, G.J., Bai, Y., Lee, J.-M., Godino, J., Saint-Maurice, P.F., & Carr, L. (2019). Standardizing analytic methods and reporting in activity monitor validation studies. *Medicine & Science in Sports & Exercise*, *51*(8), 1767. PubMed ID: 30913159 doi:10.1249/MSS.0000000000001966
- Woodman, J.A., Crouter, S.E., Bassett, D.R., Jr., Fitzhugh, E.C., & Boyer, W.R. (2017). Accuracy of consumer monitors for estimating energy expenditure and activity type. *Medicine & Science in Sports & Exercise*, *49*(2), 371–377. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip.uid&db=sph&AN=120904259&site=ehost-live&scope=site>
- Yavelberg, L., Zaharieva, D., Cinar, A., Riddell, M.C., & Jamnik, V. (2018). A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: A novel approach. *Journal of Diabetes Science and Technology*, *12*(3), 569–576. PubMed ID: 29320885 doi:10.1177/1932296817750401
- Zhang, X., McClean, D., Ko, E., Morgan, M.A., & Schmitz, K. (2017). Exercise among women with ovarian cancer: A feasibility and pre-/post-test exploratory pilot study. *Oncology Nursing Forum*, *44*(3), 366–374. PubMed ID: 28635971 doi:10.1188/17.ONF.366-374