

Review

# Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review

Daniel Fuller<sup>1,2,3\*</sup>, PhD; Emily Colwell<sup>1\*</sup>, MSc; Jonathan Low<sup>1\*</sup>, MSc; Kassia Orychock<sup>1</sup>, MSc; Melissa Ann Tobin<sup>4</sup>, MSc; Bo Simango<sup>1</sup>, MSc; Richard Buote<sup>4</sup>, MSc; Desiree Van Heerden<sup>5</sup>; Hui Luan<sup>6</sup>, PhD; Kimberley Cullen<sup>1,3</sup>, PhD; Logan Slade<sup>4</sup>, BSc; Nathan G A Taylor<sup>7</sup>, MPH

<sup>1</sup>School of Human Kinetics and Recreation, Memorial University, St. John's, NL, Canada

<sup>2</sup>Department of Computer Science, Memorial University, St. John's, NL, Canada

<sup>3</sup>Division of Community Health and Humanities, Faculty of Medicine, Memorial University, St. John's, NL, Canada

<sup>4</sup>Faculty of Medicine, Memorial University, St. John's, NL, Canada

<sup>5</sup>Faculty of Engineering, Memorial University, St. John's, NL, Canada

<sup>6</sup>Department of Geography, University of Oregon, Eugene, OR, United States

<sup>7</sup>School of Health Administration, Dalhousie University, Halifax, NS, Canada

\*these authors contributed equally

**Corresponding Author:**

Daniel Fuller, PhD  
School of Human Kinetics and Recreation  
Memorial University of Newfoundland  
St. John's, NL, A1C 5S7  
Canada  
Phone: 1 7098647270  
Email: [dfuller@mun.ca](mailto:dfuller@mun.ca)

## Abstract

**Background:** Consumer-wearable activity trackers are small electronic devices that record fitness and health-related measures.

**Objective:** The purpose of this systematic review was to examine the validity and reliability of commercial wearables in measuring step count, heart rate, and energy expenditure.

**Methods:** We identified devices to be included in the review. Database searches were conducted in PubMed, Embase, and SPORTDiscus, and only articles published in the English language up to May 2019 were considered. Studies were excluded if they did not identify the device used and if they did not examine the validity or reliability of the device. Studies involving the general population and all special populations were included. We operationalized validity as criterion validity (as compared with other measures) and construct validity (degree to which the device is measuring what it claims). Reliability measures focused on intradevice and interdevice reliability.

**Results:** We included 158 publications examining nine different commercial wearable device brands. Fitbit was by far the most studied brand. In laboratory-based settings, Fitbit, Apple Watch, and Samsung appeared to measure steps accurately. Heart rate measurement was more variable, with Apple Watch and Garmin being the most accurate and Fitbit tending toward underestimation. For energy expenditure, no brand was accurate. We also examined validity between devices within a specific brand.

**Conclusions:** Commercial wearable devices are accurate for measuring steps and heart rate in laboratory-based settings, but this varies by the manufacturer and device type. Devices are constantly being upgraded and redesigned to new models, suggesting the need for more current reviews and research.

(*JMIR Mhealth Uhealth* 2020;8(9):e18694) doi: [10.2196/18694](https://doi.org/10.2196/18694)

**KEYWORDS**

commercial wearable devices; systematic review; heart rate; energy expenditure; step count; Fitbit; Apple Watch; Garmin; Polar

## Introduction

Globally, physical inactivity is a pressing public health concern. A recent report suggested that about 23% of adults and 81% of school-going adolescents are not meeting physical activity guidelines [1]. Government organizations have attempted to improve these numbers by implementing initiatives aimed at promoting physical activity. Though the successful promotion of physical activity is a complex multifaceted issue, behavior change is a well-established method to increase physical activity [2]. Metrics defining physical activity guidelines from commercial wearable devices have been developed, including 10,000 steps per day [3,4] and 100 steps per minute for moderate to vigorous activity [5]. However, research has shown variation in step count among devices, and the applicability of these metrics may vary by device brand and device type [6].

Research examining consumer wearable devices, such as watches, pendants, armbands, and other accessories, is associated with various labels including Quantified Self [7] and mobile health (mHealth) [8]. These consumer wearable devices are becoming increasingly popular for purchase and use. It has been estimated that in the year 2019, 225 million consumer wearables were sold [9], and studies have suggested that more than a third of adults in Canada and Australia own and use a consumer wearable device [10,11]. Despite their popularity, research is equivocal about whether commercial wearable devices are valid and reliable methods for estimating metrics associated with physical activity including steps, heart rate, and energy expenditure.

In a recent review of 10 articles, Bunn et al [12] noted tendencies of wearables to underestimate energy expenditure, heart rate, and step count. Fitbit wearables were highly correlated with criterion measures of step count during laboratory-based assessment and had consistently high interdevice reliability for both step count and energy expenditure [13]. However, this review found that these devices tended to underestimate energy expenditure, which is consistent with a separate review of Fitbit accuracy [14] indicating that Fitbit wearables provide accurate measures only in limited circumstances.

Commercial wearable devices have the potential to allow for population-level measurement of physical activity and large-scale behavior change. However, questions remain about their reliability and validity. This is especially true of smaller and newer manufacturers of wearable devices for which few or no reliability and validity studies have been conducted. The purpose of this systematic review was to outline and summarize information about the validity and reliability of wearables in measuring step count, heart rate, and energy expenditure in any population. The information summarized herein can be used to inform consumers and can aid researchers in study design when selecting physical activity monitoring devices.

## Methods

### Design

This systematic review was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and

Meta-Analyses (PRISMA) guidelines [14]. The review was not registered with PROSPERO. Full-length peer-reviewed original research articles, short reports, and letters to the editor published from January 1, 2000, through May 28, 2019, were included in the search. We limited the search to articles published after the year 2000 because commercial wearable devices were not truly available before that time.

### Search Strategy

We conducted a literature search of the following databases: MEDLINE via PubMed (1946 to present); Embase (1947 to present); and SPORTDiscus with full text (1920 to present) via EBSCO. The reference lists of eligible papers were reviewed for additional pertinent references.

A librarian (KR) developed the MEDLINE search strategy, which was peer reviewed by a second librarian according to the Peer Review of Electronic Search Strategies (PRESS) 2015 Guideline Statement [15]. The MEDLINE strategy, which included Medical Subject Heading terms and text words, was translated for the other databases using database-specific controlled vocabulary. We searched the literature using multiple combinations and forms of the following key terms: *accelerometer, fitness tracker, activity monitor, step count, wearable device, validity, reliability, accuracy, Fitbit, Garmin, Misfit, Jawbone, UnderArmour, Samsung, Apple watch, GENEactiv, Empatica, Mio, Amiigo, Xiaomi, Actigraph, Withings, and Sensewear* (see [Multimedia Appendix 1](#) for the full search strategies). An English language limit was applied. We included any abstracts and conference proceedings, as well as articles examining any population in the initial search. References were imported into EndNote X8 software (Clarivate Analytics) where duplicate references were removed. The remaining references were then imported into Covidence software (Veritas Health Innovation) for screening.

### Study Selection Strategy

The web-based systematic review software Covidence was used for this review. The titles and abstracts of the studies included from the initial database search were independently assessed by at least two authors from the team. Conflicts arising during any step of the screening for inclusion/exclusion were resolved by a third author or by consensus. Following the title and abstract screening, full-text documents of the selected studies were searched and retrieved and were independently assessed for inclusion by at least two authors (EC, JL, and DF). Any conflicts were resolved by discussion and consensus. All reviewers strictly adhered to the defined inclusion criteria.

### Eligibility Criteria

Studies that met the following criteria were included in the review: (1) use of any consumer-wearable model from the brand Apple Inc, Empatica, Fitbit, Garmin, Jawbone, Mio, Misfit, Polar, Samsung, UnderArmour, Withings, or Xiaomi; (2) specific examination of the reliability and validity measures of the aforementioned brands; and (3) examination of the device's ability to measure a variable (step count, heart rate, or energy expenditure). Studies with fewer than 10 participants were excluded, as has been done in previous work [13]. Validity of the wearable devices was defined as follows [16]:

- Criterion validity: comparing the devices to a criterion measure of steps, heart rate, or energy expenditure.

Reliability of the trackers included the following [16]:

- Intradvice reliability: consistent test-retest results conducted within the same device.
- Interdevice reliability: consistent results across the same model of wearable device measured at the same time and worn at the same location.

The main exclusion criteria were non-English studies, opinion/magazine articles, and systematic reviews. The initial database search and title/abstract screening included articles examining the accuracy of research-grade wearable devices, but the number of returned results was unmanageable. In order to further elucidate the research question in regard to consumer-wearable devices, before full-text screening, the decision was made to exclude all studies examining the reliability and validity of research-grade devices (Actigraph, GENEactiv, Amiigo, Sensewear Armband, Yamax, Omron, Kenze Lifecorder, Digiwalker, Actical, and Actiheart). Studies in which heart rate and energy expenditure estimates were collected using a chest strap heart rate monitor and transmitted to a wearable device were also excluded. Following text screening, the decision was made to exclude abstracts and conference papers. Following data extraction, the decision was made to exclude all studies examining Jawbone commercial wearables, as the company's application program interface (API) was taken offline in 2018, rendering associated devices defunct. Studies were included in the final review if they had extractable data for the following criterion validity measures: correlation coefficient, group mean or percentage difference, median or mean absolute percentage error (MAPE), or level-of-agreement analysis, or had correlation coefficients for reliability measures. Authors were not contacted if these data were not reported in published or supplementary material. The remaining articles were those that met the inclusion criteria (consumer-grade wearables).

### Risk of Bias

In our risk of bias assessment, comparisons that did not report group percentage differences or correlation coefficients (n=192) were excluded from the quantitative analysis. However, rather than exclude these comparisons and studies from the review completely, we included them in a narrative summary of how the measures reported were or were not consistent with exploration of percentage measurement error and correlation.

### Data Extraction

We first conducted and documented an in-depth web search of the available consumer-wearable models and their specifications (placement, size, weight, cost, and connectivity). The data extraction process then consisted of the following: (1) categorizing the selected full-text articles into reliability or validity studies (EC, JL, and DF); (2) using a modification of the modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) validation subscale used by Feehan et al [13] and an a priori modified COSMIN reliability subscale (Multimedia Appendix 2) to assess the quality and risk of bias of each study (EC and DF); (3)

extracting the key characteristics from each selected publication and compiling them into tables. Details from each reviewer were compared, and inconsistencies were resolved through consensus before compiling the results (EC and DF).

Data extracted included characteristics of studies, participants, and devices, including study setting and activity type, outcomes measured, and type of criterion measure used. Correlation coefficients were extracted for all reliability comparisons reported in each study. Correlation coefficients, percentage difference and group mean values, MAPE values, and level-of-agreement data were extracted for all validity comparisons where available. Where group percentage differences were not reported, we calculated group percentage error ( $[\text{wearable}_{\text{mean}} - \text{criterion}_{\text{mean}}] / \text{criterion}_{\text{mean}} \times 100$ ) to allow for comparison across studies. We split a small number of studies (n=10) into "substudies" (n=21), where separate populations were examined in the same publication (see Multimedia Appendix 3 for a more detailed breakdown).

### Syntheses

Given the wide range of testing conditions and reported outcomes, we were unable to conduct meta-analyses of the extracted data. We instead conducted a narrative synthesis of the available quantitative data within each examined measure (step count, heart rate, and energy expenditure) using correlation comparisons and group percentage difference as the common metrics for criterion validity and correlation coefficient as the common metric for reliability.

Our interpretation of measurement accuracy was focused on acceptable limits of percentage difference of  $\pm 3\%$  in controlled settings and percentage difference of  $\pm 10\%$  in free-living settings, as outlined in previous work [13]. We interpreted correlation coefficients as follows: 0 to  $<0.2$ , very weak;  $\geq 0.2$  to  $<0.4$ , weak;  $\geq 0.4$  to  $<0.6$ , moderate;  $\geq 0.6$  to  $<0.8$ , strong; and  $\geq 0.8$  to 1.0, very strong [17]. We completed all quantitative analyses and plots using RStudio version 1.2.1335 (RStudio Inc) and R version 3.6.0 (The R Foundation).

Secondary analyses explored device brand. Brands were only included in these analyses when the group had 10 or more comparisons available for the measure. Studies that did not report data allowing for the examination of group percentage measurement error were still included in the review if they reported level of agreement or MAPE data. Such studies were included in the risk of bias assessment, the synthesis of study characteristics, and the narrative synthesis of study results.

### Availability of Data and Materials

Data are publicly available on the BeapLab Dataverse [18], and the analysis code is available on Github [19].

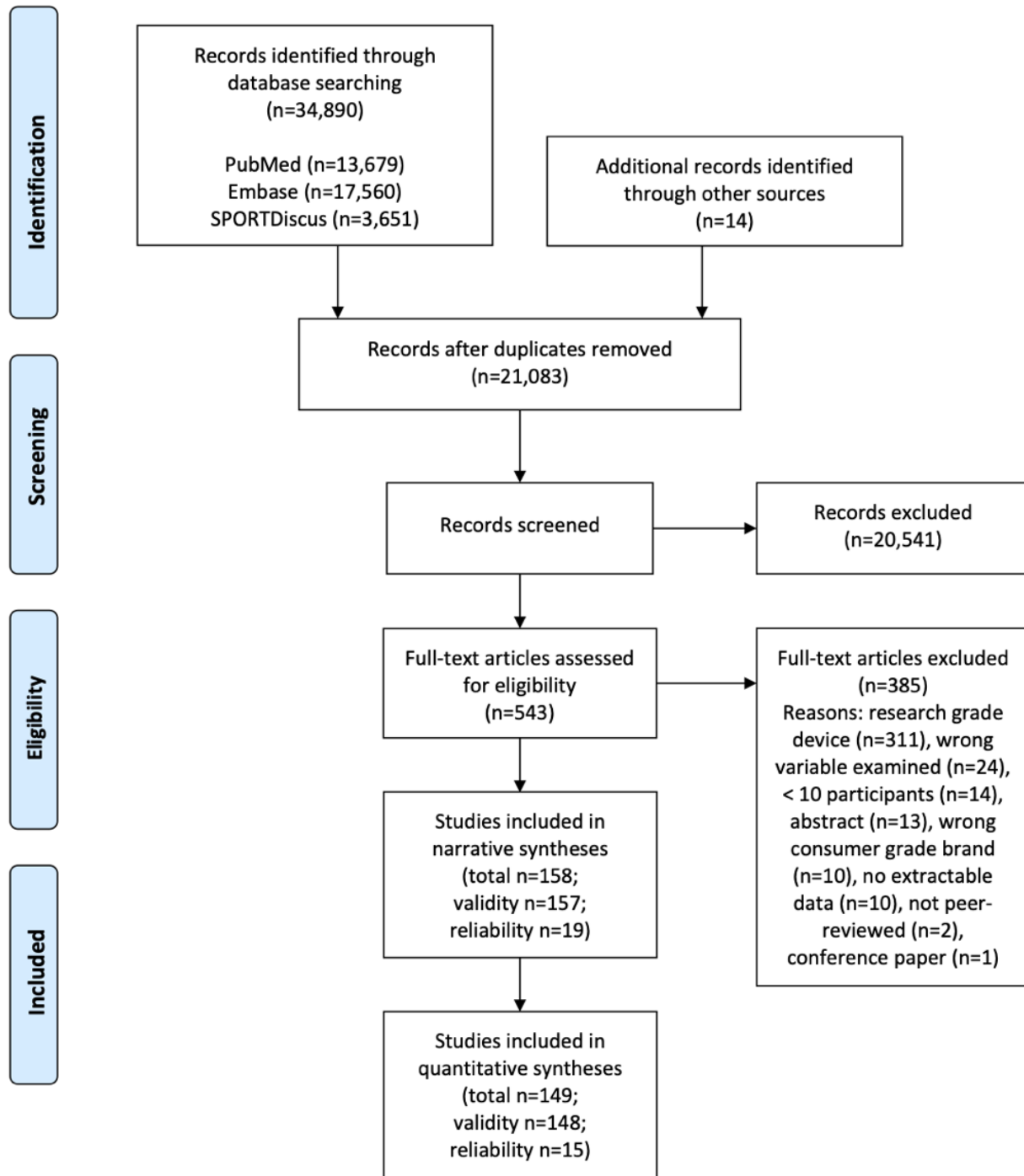
## Results

The initial literature search from the three databases yielded 34,890 unique citations (13,679 [39.21%] from PubMed, 17,560 [50.33%] from Embase, and 3651 [10.46%] from SPORTDiscus). Fourteen additional records were identified through other sources (eg, article reference lists and social media). After duplicate references were removed, 21,083

citations remained. Based on the subsequent title and abstract screening, 20,541 were rejected because they did not meet the inclusion criteria or met the exclusion criteria. Of the 542 that remained for full-text screening, 385 (71.0%) were further excluded for the following reasons: research-grade devices (n=311, 57.4%), wrong variable examined (n=24, 4.4%), fewer than 10 participants (n=14, 2.6%), abstracts (n=13, 2.4%), wrong consumer-grade brand examined (n=10, 1.9%); devices were

Yamax, Omron, Kenz Lifecorder, Digiwalker, and uniaxial Actical/Actiheart), no extractable data (n=10, 1.9%), not peer reviewed (n=2, 0.4%), and conference paper (n=1, 0.2%). As a result, a total of 158 publications were included in this systematic review (Figure 1) [14]. Table 1 shows the details of the device brand, model, year, and status (current model or discontinued) in the included studies.

Figure 1. PRISMA flow chart for systematic review of the reliability and validity of commercial wearable devices.



**Table 1.** Device brand, model, year, current status, wear location, and studies used for the current systematic review.

Brand	Model	Year	Status	Wear location	Studies
Apple	Watch	2015	Discontinued	Wrist	[20-40]
Apple	Watch Series 2	2016	Discontinued	Wrist	[41-44]
Fitbit	Alta	2016	Current model	Wrist	[45]
Fitbit	Blaze	2016	Discontinued	Wrist	[22,40,43]
Fitbit	Charge	2014	Discontinued	Wrist	[45-56]
Fitbit	Charge 2	2016	Discontinued	Wrist	[23,30,43,44,57-63]
Fitbit	Charge HR	2015	Discontinued	Wrist	[20,21,29,32,34,36,38,45,53,64-82]
Fitbit	Classic	2009	Discontinued	Ankle/foot or waist/hip	[83-87]
Fitbit	Flex	2013	Discontinued	Thigh or wrist	[45,50,72,79,80,88-117]
Fitbit	Flex 2	2017	Current model	Wrist	[113]
Fitbit	Force	2013	Discontinued	Wrist	[118,119]
Fitbit	One	2012	Discontinued	Ankle/foot, pant pocket, waist/hip, or wrist	[34,49,52,73,80,88,90,92,93,98,100,102,103,110,116-118,120-138]
Fitbit	Surge	2015	Discontinued	Wrist	[27,35,42,45,54,82,139-143]
Fitbit	Ultra	2011	Discontinued	Chest, pant pocket, upper arm, waist/hip, or wrist	[85,144-148]
Fitbit	Zip	2012	Current model	Ankle/foot, pant pocket, shin, or waist/hip	[34,45,46,51,88,89,92,93,96,103,112,119,127,129,131,141,149-161]
Garmin	Fenix 3 HR	2016	Discontinued	Wrist	[41]
Garmin	Forerunner 225	2015	Discontinued	Wrist	[21,162]
Garmin	Forerunner 235	2015	Current model	Wrist	[40,139,163]
Garmin	Forerunner 405CX	2009	Discontinued	Wrist	[164]
Garmin	Forerunner 735XT	2016	Current model	Wrist	[35]
Garmin	Forerunner 920XT	2014	Discontinued	Wrist	[53]
Garmin	Vivoactive	2015	Discontinued	Wrist	[53]
Garmin	Vivofit	2014	Discontinued	Wrist	[35,46,50,52,53,89,92,104,114,122,130,143,150,159,165-169]
Garmin	Vivofit 2	2015	Discontinued	Wrist	[34,127,170]
Garmin	Vivofit 3	2016	Discontinued	Wrist	[168,171]
Garmin	Vivosmart	2014	Discontinued	Wrist	[32,53,75]
Garmin	Vivosmart HR	2015	Discontinued	Wrist	[36,43,65]
Garmin	Vivosmart HR+	2016	Current model	Wrist	[58,63,140]
Mio	Alpha	2013	Discontinued	Wrist	[25,38,71]
Mio	Fuse	2015	Discontinued	Wrist	[54,64]
Misfit	Flash	2015	Discontinued	Waist/hip	[32]
Misfit	Shine	2012	Discontinued	Ankle/foot, chest, pant pocket, waist/hip, or wrist	[74,79,89,96,99,104,105,131,159,169]



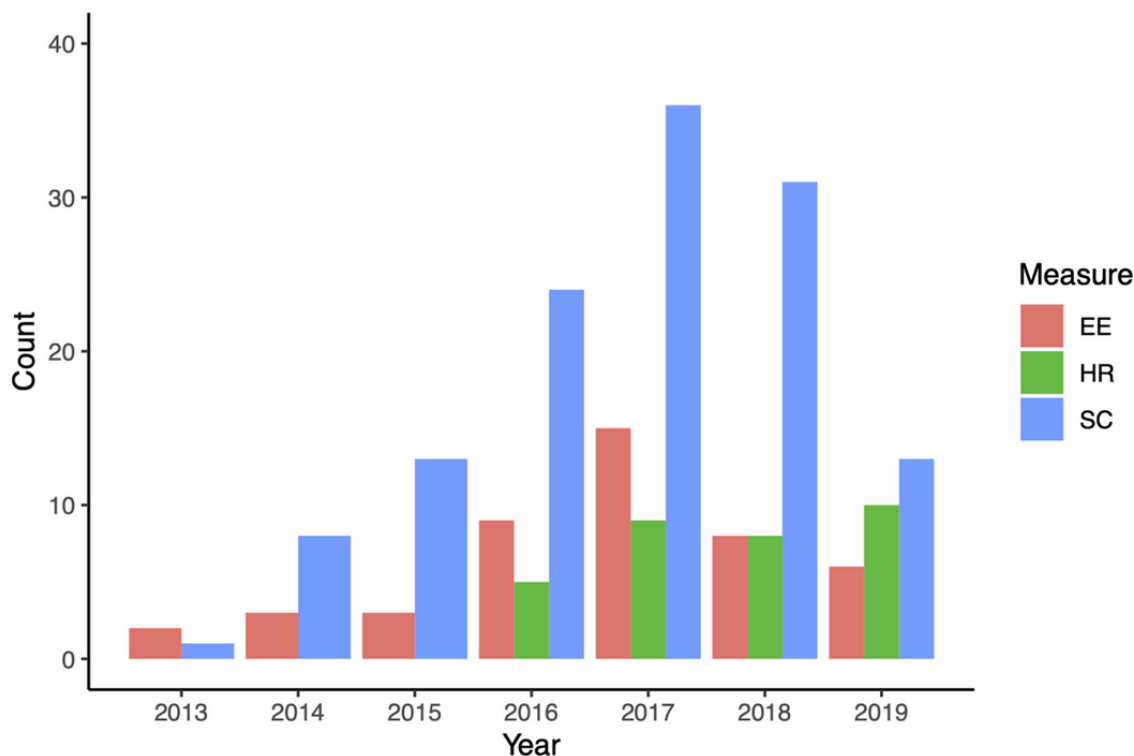
Brand	Model	Year	Status	Wear location	Studies
Polar	A300	2015	Discontinued	Wrist	[172]
Polar	A360	2015	Discontinued	Wrist	[25,43,140]
Polar	Active	2011	Discontinued	Wrist	[173]
Polar	Loop	2013	Discontinued	Wrist	[32,50,53,79,89,167]
Polar	M600	2016	Current	Wrist	[56]
Polar	V800	2016	Discontinued	Wrist	[174]
Samsung	Gear 2	2014	Discontinued	Wrist	[140]
Samsung	Gear S	2014	Discontinued	Wrist	[32,38]
Samsung	Gear S2	2015	Discontinued	Wrist	[35]
Samsung	Gear S3	2016	Discontinued	Wrist	[42,44]
Withings	Pulse	2013	Discontinued	Collar, pant pocket, waist/hip, or wrist	[89,96,131,166]
Withings	Pulse O2	2013	Discontinued	Collar, waist/hip, or wrist	[104,122,123,169,175]
Withings	Pulse Ox	2014	Current model	Waist/hip or wrist	[53,58]
Xiaomi	Mi Band	2014	Discontinued	Wrist	[42]
Xiaomi	Mi Band 2	2016	Discontinued	Wrist	[81]

### Study and Participant Characteristics

Of the 158 publications included, 143 were full-text research articles, 10 were brief reports, and five were letters to the editor. Publication year ranged from 2013 to 2019, with the amount of publications increasing from 2013 to 2017 (2013, n=2; 2014, n=8; 2015, n=11; 2016, n=30; 2017, n=43). We also included an additional 40 and 24 studies published in 2018 and 2019, respectively.

Within those 158 publications, 169 studies/substudies were identified. Among these, 168 (99.4%) examined validity and

19 (11.2%) examined reliability. Moreover, 126 studies examined step count (125 validity and 16 reliability), 32 examined heart rate (32 validity and 3 reliability), and 43 examined energy expenditure (42 validity and 5 reliability) (Figure 2). Furthermore, 130 examined populations in a controlled environment and 48 examined populations in a free-living environment. A total of 1838 comparisons were identified, of which 166 examined reliability (mean 8, SD 11 per reliability study; range 1-40) and 1672 examined validity (mean 10, SD 15 per validity study; range 1-98).

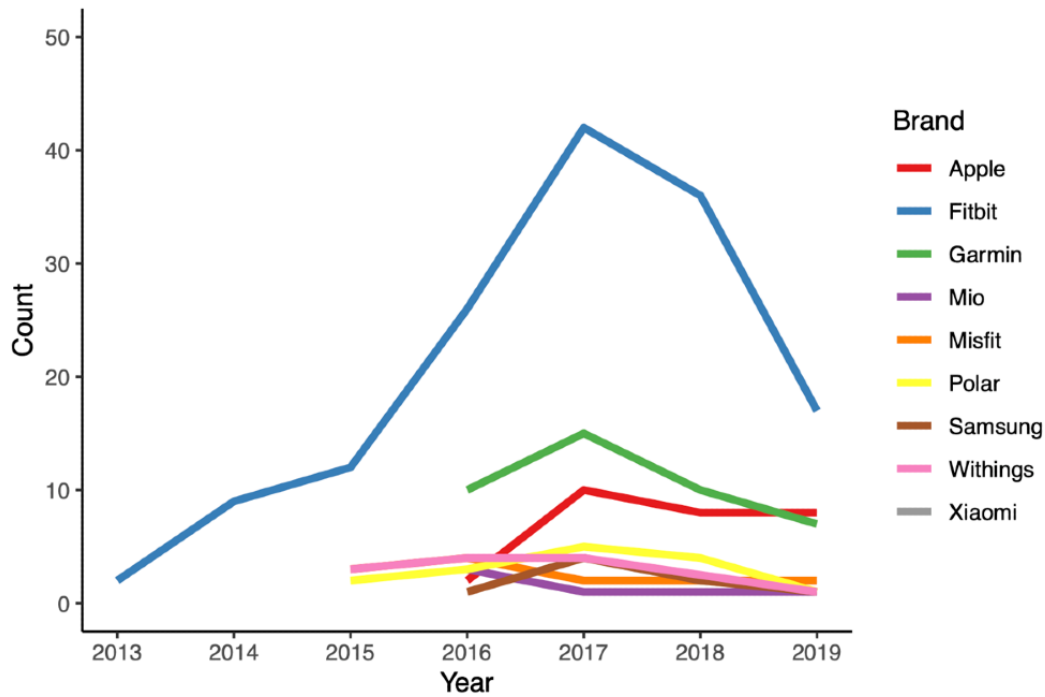
**Figure 2.** Number of studies published per year by measurement type. EE: energy expenditure; HR: heart rate; SC: step count.

The 169 studies/substudies comprised a total of 5934 participants, with a mean of 35 (SD 27) participants per study (range 10-185). One hundred and sixty-one studies reported sex, and 51.08% (2861/5601) of participants were female. One hundred and fifty-eight studies reported age, with a mean participant age of 36.8 years (SD 18.3; range 3.7-87 years). One hundred and fifty-nine studies examined adult populations (age  $\geq 18$  years) and 10 studies examined children. One hundred and thirty-three studies included only healthy participants, while the other 36 studies included participants with mobility limitations and/or chronic diseases (Multimedia Appendix 4).

Fitbit consumer-grade wearables were examined most frequently (144 studies examining 12 models), followed by Garmin (42

studies, 13 models), Apple (28 studies, 2 models), Polar (15 studies, 6 models), Misfit (13 studies, 2 models), Withings (12 studies, 2 models), Samsung (8 studies, 4 models), Mio (6 studies, 2 models), and Xiaomi (2 studies, 2 models) (a complete list of examined models is provided in Multimedia Appendix 5) (Figure 3). Wearables were typically examined while worn on the wrist ( $n=131$ , examining at least one wrist-worn device) or at the waist/hip ( $n=71$ , locations included the waist, hip, belt, and pants pocket). Substantially fewer studies examined wearables worn on the torso ( $n=14$ , locations included the chest, bra, lanyard, and shirt collar) and lower limb ( $n=13$ , locations included the thigh, shin, ankle, and foot).

**Figure 3.** Line graph of studies published per year by device brand.

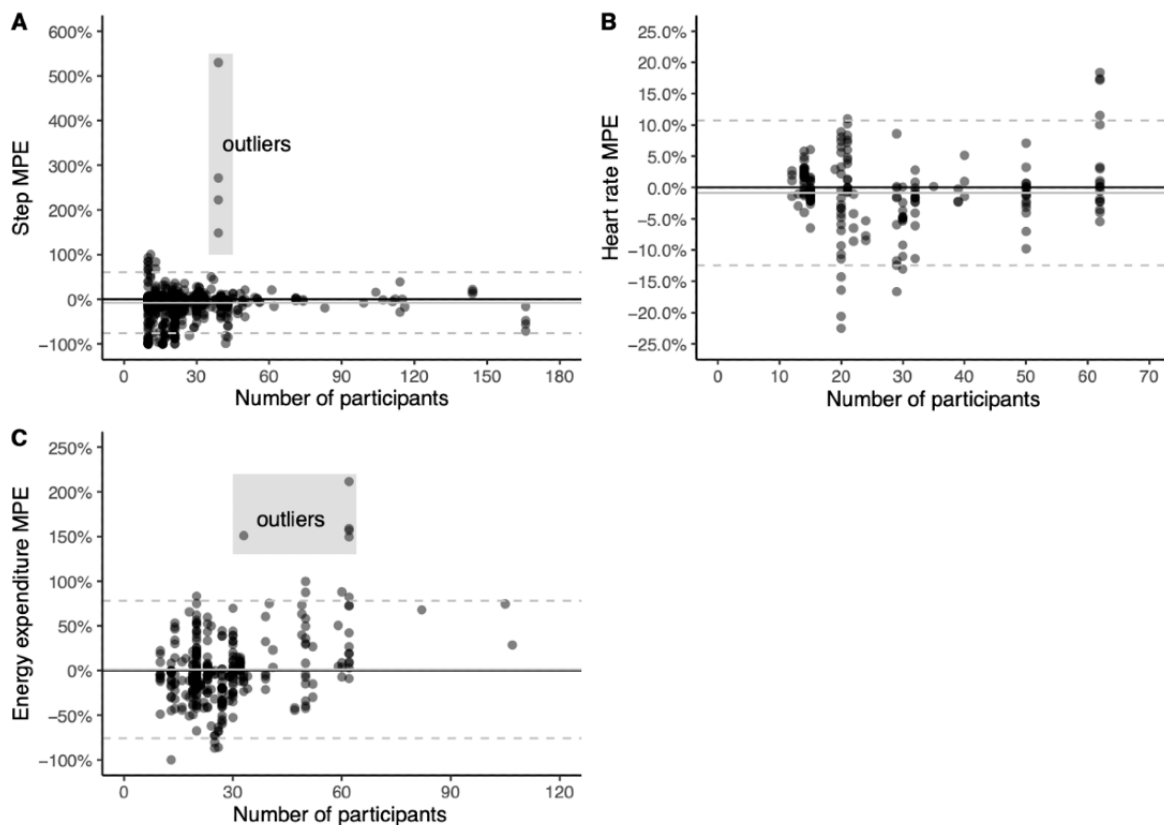


**Risk of Bias**

Of 169 studies, 140 (82.8%; 1640 of 1838 [89.23%] comparisons) were rated fair or poor for sample size (<50 participants), but were not excluded from the analysis owing to the paucity of studies with excellent ( $\geq 100$  participants,  $n=7$ )

and good (50-99 participants,  $n=22$ ) sample sizes. We additionally explored the potential for bias related to sample size in step count, heart rate, and energy expenditure by examining the percentage error dispersion by sample size using scatter plots (Figure 4).

**Figure 4.** Mean percentage error (MPE) plots by study sample size for step count, heart rate, and energy expenditure. The solid black line represents zero. The solid grey line represents average MPE for all data points. The dashed grey lines represent the 95% CIs.





In these examinations, we saw no apparent systematic bias for measurement error beyond a small number of comparisons showing extreme overestimation (four comparisons in step count and five comparisons in energy expenditure). The four extreme outliers for step count involved measurement during sedentary and light physical activity in a single study with fewer than 40 participants [20] and were likely inflated by the limited number of steps accumulated during those bouts. As a result, we excluded these four comparisons from the quantitative syntheses. Upon closer examination of the five extreme outliers for energy expenditure (four occurred in a study with greater than 60 participants [21] and one occurred in a study with fewer than 40 participants [41]), we determined that these were likely true reflections of tendencies to overestimate energy expenditure during sedentary and low-intensity activities, and therefore, we included these five comparisons in the quantitative syntheses.

### Validity: Controlled Settings

We examined criterion validity for step count, heart rate, and energy expenditure separately for controlled and free-living settings. For controlled settings, we also had sufficient data to examine validity by brand and devices within brands.

#### Validity for Step Count in Controlled Settings

A total of 90 studies (979 comparisons) examined wearable device step count measurements compared with reference standard criterion measures of manual counting [32,34-38,42,46,47,50-53,57,58,72,80-84,88-102,109,114-125,138-141,144-147,149-153,158-161,165,169-171,173] and accelerometry [20,60,64-66,85,103,109,126-128,148,154,164] (Multimedia Appendix 6). Of these, 67 studies recruited healthy adults (mean age 35.4 years, SD 17.4 years), 20 studies recruited adults living with limited mobility/chronic diseases (mean age 60.1 years, SD 10.5 years), two studies recruited children living with limited mobility/chronic diseases (mean age 12.5 years, SD 2.9 years), and one study recruited healthy children (mean age 3.7 years, SD 0.6 years). Wearable devices were worn on the lower limb (foot, ankle, shin, and thigh), torso, waist/hip, and wrist.

Group measurement error was reported or calculable for 805 of the 979 comparisons, regardless of the criterion measure. Of these, 45.2% (n=364) were within  $\pm 3\%$  measurement error, 42.7% (n=344) were below  $-3\%$  measurement error, and 12.1% (n=97) were above 3% measurement error. The overall tendency was to underestimate step count (mean:  $-9\%$ , median:  $-2\%$ ).

#### Validity for Heart Rate in Controlled Settings

A total of 29 studies (266 comparisons) examined wearable device heart rate measurements compared with reference

standard criterion measures, including electrocardiography [22,23,38-40,43,44,54,61,62,67-70,142,162,176], Polar brand chest straps [20,21,24-28,58,63,71,163], and pulse oximetry [66], in controlled settings (a detailed list of the criterion measures used is presented in Multimedia Appendix 6). Of these, 24 studies recruited healthy adults (mean age 29.8 years, SD 10.5 years), four studies recruited adults living with limited mobility/chronic diseases (mean age 59.6 years, SD 9.0 years), and one study recruited children undergoing surgery (mean age 8.2 years, SD 3.1 years). All wearable devices were worn on the wrist.

Group measurement error was reported or calculable for 177 of 266 comparisons, regardless of the criterion measure. Of these, 56.5% (n=100) were within  $\pm 3\%$  measurement error, 24.9% (n=44) were below  $-3\%$  measurement error, and 18.6% (n=33) were above 3% measurement error. There was a slight overall tendency toward underestimation of heart rate (estimated median error:  $-1\%$ ).

#### Validity for Energy Expenditure in Controlled Settings

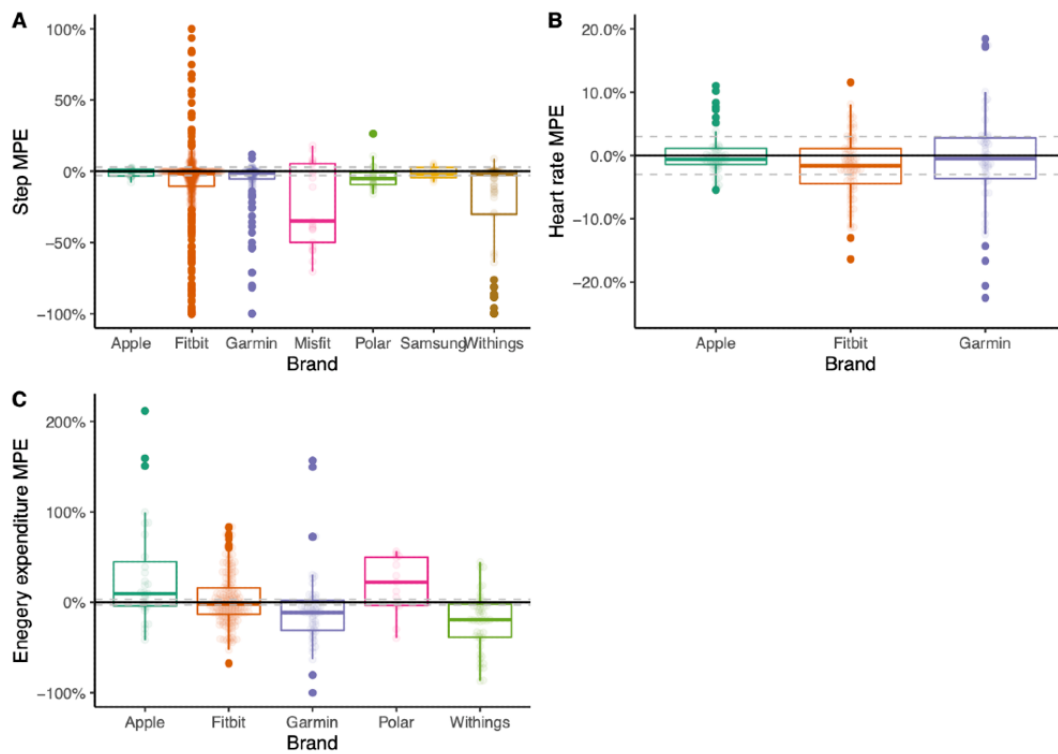
A total of 36 studies (312 comparisons) examined wearable device energy expenditure measurements compared with reference standard criterion measures, including direct calorimetry [86,104] and indirect calorimetry [20,21,29-31,38,39,41-43,53,55,63,66,73,85,87,93,95,97,103,105,116,117,129,130,142,143,146,148,159,165,166,177], in controlled settings. Of these, 35 studies recruited healthy adults (mean age 27.2 years, SD 7.1 years), and one study recruited adults living with cardiovascular disease (mean age 64.2 years, SD 2.3 years). Wearable devices were worn on the wrist, waist/hip, and torso.

Group measurement error was reported or calculable for 305 of the 312 comparisons, regardless of the criterion measure. Of these, 9.2% (n=28) were within  $\pm 3\%$  measurement error, 54.1% (n=165) were below  $-3\%$  measurement error, and 36.7% (n=112) were above 3% measurement error. Studies showed a tendency to underestimate energy expenditure and to provide inaccurate measures of energy expenditure compared with the criterion.

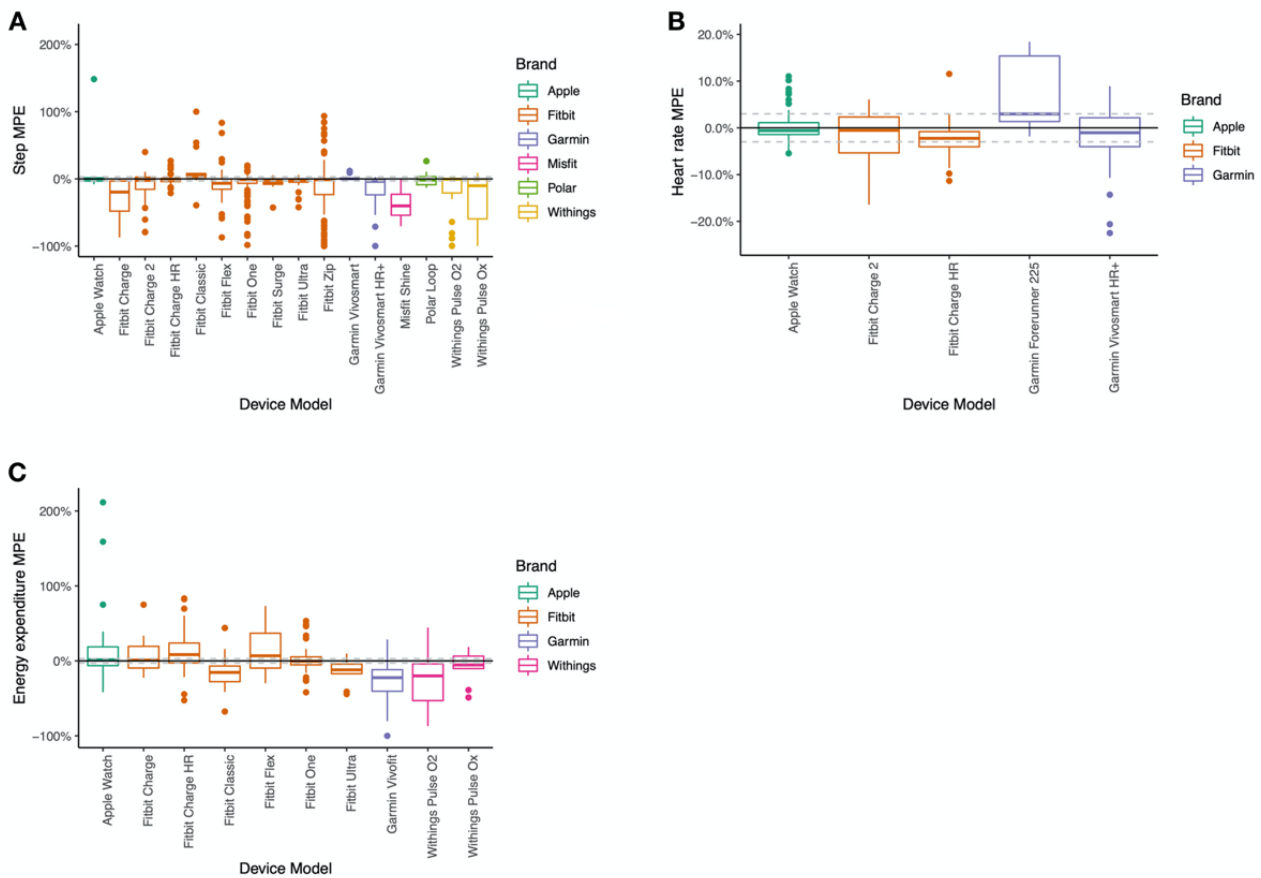
#### Validity in Controlled Settings by Brand

Figure 5 shows the mean percentage error (MPE) for step count, heart rate, and energy expenditure by device brand for devices with 10 or more comparisons. Figure 6 shows the MPE for step count, heart rate, and energy expenditure by device brand and model for devices with 10 or more comparisons.

**Figure 5.** Box plots representing mean percentage error (MPE) for steps, heart rate, and energy expenditure by device brand for devices with 10 or more comparisons.



**Figure 6.** Box plots representing mean percentage error (MPE) for steps, heart rate, and energy expenditure by device brand and model for devices with 10 or more comparisons.



### **Validity for Step Count by Brand**

We observed that the error level varied by device brand (Figure 5). Withings and Misfit wearables consistently underestimated step count, and Apple and Samsung had less measurement variability than other brands. There are possible interactions between the number and size of studies and device wear location that may influence the brand comparisons. For example, Apple Watch and Samsung have the tightest ranges for step count estimates but have relatively fewer studies compared with other brands.

### **Validity for Heart Rate by Brand**

For heart rate, measurement error also varied by device brand (Figure 5). Apple Watch was within  $\pm 3\%$  71% (35/49) of the time, while Fitbit wearables were within  $\pm 3\%$  51% (36/71) of the time and Garmin wearables were within  $\pm 3\%$  49% (23/47) of the time. Despite similar  $\pm 3\%$  measurement error rates, Fitbit appeared to underestimate heart rate more than Apple Watch and Garmin.

### **Validity for Energy Expenditure by Brand**

For energy expenditure estimates, no brand of wearable was within  $\pm 3\%$  measurement error more than 13% of the time (Figure 5). Underestimation of energy expenditure (less than  $-3\%$ ) was observed in Garmin wearables 69% (37/51) of the time and in Withings wearables 74% (34/46) of the time. Conversely, Apple wearables overestimated energy expenditure 58% (18/31) of the time and Polar wearables overestimated energy expenditure 69% (9/13) of the time. Fitbit devices tended to provide inaccurate measures compared with the criterion, underestimating 48.4% (76/157) of the time and overestimating 39.5% (62/157) of the time, despite the boxplot in Figure 5 showing a reasonable median value for accuracy.

### **Validity: Free-Living Settings**

There were relatively few studies on wearable device validity in free-living conditions. Fitbit was the only brand with more than 10 studies published for step count validity in free-living conditions, and no brands had more than 10 studies for heart rate or energy expenditure. As a result, we have not shown plots of MPE for free-living conditions.

### **Validity for Step Count in Free-Living Settings**

A total of 42 studies (84 comparisons) examined wearable device step count measurements compared with the reference standard criterion measure of accelerometry [33,45,48,49,56,59,60,64,74-76,89,96,101,106-112,120,131-136,149,154-156,159,167,168,172-175] in free-living settings (Multimedia Appendix 6). Of these, 28 studies recruited healthy adults (mean age 33.7 years, SD 13.9 years), nine studies recruited adults living with limited mobility/chronic diseases (mean age 60.1 years, SD 11.2 years), four studies recruited healthy children (mean age 12.5 years, SD 2.6 years), and one study recruited children living with cardiac diseases (mean age 13 years, SD 2.2 years). Wearable devices were worn on the lower limb (foot, ankle, and shin), torso, waist/hip, and wrist.

Group measurement error was reported or calculable for 69 of the 84 comparisons, regardless of the criterion measure. Of these, 42% (n=29) were within  $\pm 10\%$  measurement error, 17%

(n=12) were below  $-10\%$  measurement error, and 41% (n=28) were above 10% measurement error. The overall tendency was slight overestimation of step count (mean: 5%, median: 6%). Among the remaining comparisons, 11 of 15 reported MAPE, of which 40% (n=6) were below 10% measurement error and 60% (n=9) were above 10% measurement error.

### **Validity for Heart Rate in Free-Living Settings**

Three studies (five comparisons) examined wearable device heart rate compared with the reference standard criterion measure of a Polar brand chest strap in free-living settings [75,77,78]. Of these, one study recruited healthy adults (mean age 25.4 years, SD 3.7 years), one study recruited healthy children (mean age 8 years, SD 1.8 years), and one study recruited adults recovering from stroke (mean age 64.4 years, SD 15 years). All wearable devices were worn on the wrist. Group measurement error was reported or calculable for one of the five comparisons, with the Fitbit Charge HR falling within  $\pm 10\%$  measurement error in the study examining healthy children. Three of the four remaining comparisons examined the Fitbit Charge HR in adults and noted underestimation of heart rate that varied depending on activity intensity, but all reported that MAPE values fell within 10% measurement error. Correlation coefficients were strong to very strong in four of the five comparisons and moderate in one comparison examining estimation during high-intensity activity.

### **Validity for Energy Expenditure in Free-Living Settings**

Nine studies (22 comparisons) examined energy expenditure in free-living settings compared with the criterion measures of doubly labeled water [104] and accelerometry [29,49,79,101,131,172,174,175]. Eight studies recruited healthy adults (mean age 27.7 years, SD 3.8 years) and one study recruited adults with chronic obstructive pulmonary disease (mean age 66.4 years, SD 7.4 years). Wearable devices were worn on the wrist or waist/hip.

Group measurement error was reported or calculable for 17 of the 22 comparisons, regardless of the criterion measure. Of these, 18% (n=3) were within  $\pm 10\%$  measurement error, 53% (n=9) were below  $-10\%$  measurement error, and 29% (n=5) were above 10% measurement error. There was an overall tendency to underestimate energy expenditure (mean:  $-3\%$ , median:  $-11\%$ ). Xiaomi data were not analyzed in a single indirect calorimetry study owing to the lack of data [53].

### **Reliability**

Nineteen studies (166 comparisons) with sample sizes ranging from 11 [94] to 56 [151] reported inter- or intradevice reliability for Apple (seven comparisons), Fitbit (92 comparisons), Garmin (22 comparisons), Polar (one comparison), and Withings (44 comparisons). The majority of comparisons (153/166) reported interdevice reliability for step count, heart rate, or energy expenditure. No studies reported intradevice reliability for heart rate or energy expenditure. We have not reported between-brand comparisons for inter- or intradevice reliability owing to the small number of comparisons for each brand.

### ***Interdevice Reliability for Step Count***

Twelve studies (51 comparisons) with sample sizes ranging from 13 [117,138] to 56 [151] reported on interdevice reliability for step count [50,58,72,85,94,110,113,116,117,121,125,138,151,161,171]. The majority of correlation coefficients for step count interdevice reliability were very strong (n=35), with only a small number (n=3) being reported as strong.

### ***Intradvice Reliability for Step Count***

Two studies (13 comparisons) reported on intradvice reliability for step count, with sample sizes of 20 [82] and 24 [150]. Intradvice reliability correlations were very weak (n=1), weak (n=2), moderate (n=5), strong (n=2), and very strong (n=3). The mean correlation coefficient was 0.58.

### ***Interdevice Reliability for Heart Rate***

Three studies (23 comparisons) examined interdevice reliability for heart rate [24,26,58], with analyzed sample sizes ranging from 13 [24] to 21 [26]. Apple Watch showed very good interdevice reliability at 5-s epochs during treadmill bouts at 4, 7, and 10 km/h, with reliability increasing and standard typical error decreasing with increasing pace [26]. Similar standard typical error levels were seen in maximum heart rate measured during a single incremental maximal oxygen uptake test performed on a treadmill and heart rate taken from the highest 30-s mean heart rate, with somewhat lower correlation coefficients [24]. In the examination of interdevice reliability in healthy older adults, Fitbit Charge 2 showed good reliability during treadmill and overground bouts and poor reliability during hand movement tasks such as dusting [58]. During the same tasks, Garmin Vivosmart HR+ showed good reliability during all tasks and had narrower limits of agreement than Fitbit.

### ***Interdevice Reliability for Energy Expenditure***

Five studies (50 comparisons) reported on interdevice reliability [85,113,116,117,166], with analyzed sample sizes ranging from 13 [117] to 29 [113]. All five studies recruited healthy adults (mean age 26.3 years, SD 3.9 years). Correlation coefficients were reported for 16 of 50 comparisons. Of these, 13% (n=2) were rated very weak, 6% (n=1) were rated moderate, 6% (n=1) were rated strong, and 75% (n=12) were rated very strong.

## ***Discussion***

### **Overview**

The purpose of this study was to examine the validity and inter- and intradvice reliability of commercial wearable devices in measuring steps, heart rate, and energy expenditure. Our review focused on both a breadth of devices and reproducibility. Our review included nine brands and 45 devices with the number of comparisons ranging from 201 for the Fitbit Zip to one for the Garmin Forerunner 405CX and the Polar M600. For comparison, two recent reviews from 2017 included two brands and 16 devices [13] and seven brands and eight devices [79]. A review from 2016 included eight devices [32]. Along with this review, we have published our dataset and code to reproduce our findings.

Our bias assessment showed no apparent bias toward studies of different sample sizes. However, there is a strong overrepresentation of studies with 20 participants. There were some outliers in our findings; however, considering the number of included comparisons, this is to be expected.

### **Reliability and Validity**

Criterion validity of commercial wearables varied by study type (controlled or free-living), brand, and device. For step count, our review showed that in controlled laboratory settings, a higher proportion of devices showed accuracy, and this was within a tighter limit of acceptable accuracy compared with free-living conditions. In both controlled and free-living studies, when not correctly estimating steps, devices tended to underestimate values. Validity compared with criteria was the best for Apple Watch and Garmin, while the MPE values for Fitbit, Samsung, and Withings fell within  $\pm 3\%$  on average. Within brands, devices appeared to vary, with Fitbit Classic tending to overestimate steps, while Fitbit Charge tending to underestimate steps; however, the variability observed could be attributed to differences in the number of comparisons for each device and in wear locations of the devices. Our findings are consistent with previous reviews [178].

In controlled settings across all devices, heart rate was accurately measured with only a very small tendency for underestimation. Heart rate validity was only sufficiently tested in Apple Watch, Fitbit, and Garmin devices. Heart rate measured by photoplethysmography is only available in relatively new commercial wearable devices. All of the brands measured heart rate to within  $\pm 3\%$  on average in controlled settings. There were few studies examining the validity of heart rate measures in free-living conditions, but it appears that Fitbit devices may underestimate heart rate depending on activity intensity. All devices were within acceptable measurement error for heart rate. To our knowledge, this is the first systematic review to examine heart rate validity, and it appears that devices are able to measure heart rate within acceptable limits.

Energy expenditure estimates varied widely with less than 10% of estimates falling within acceptable limits in controlled settings. In many of the studies, there did appear to be a tendency for systematic over or underestimation. On average, only Fitbit measured energy expenditure to within acceptable limits, but there was wide variation around the estimate. Energy expenditure estimates also varied by model, with the Fitbit Classic underestimating the value considerably and Fitbit Charge HR overestimating the value. We hypothesize that Fitbit may provide the best, though still not acceptable, measure of energy expenditure because the algorithm employs a published equation for estimating resting metabolic rate [179]. To our knowledge, the other brands do not publish information about the energy expenditure estimates. There does not appear to be a relationship among more accurate estimates of energy expenditure in devices that include heart rate (Multimedia Appendix 7).

Interdevice reliabilities for steps, heart rate, and energy expenditure were all very strong. However, compared with validity studies, there were fewer reliability studies, and we were not able to conduct comparisons between brands or devices owing to small sample sizes. Sufficient data for intradvice



reliability was only available for step count. The results showed considerable variability within the same device for step count for Fitbit Charge HR, Fitbit Surge, Fitbit Zip, and Garmin Vivofit, with five, five, one, and two comparisons, respectively.

### Future Research

Future research in this area should focus on the following three main topics: relevance and age of the devices tested, data acquisition from the devices, and algorithms used by companies. First, relevance of the devices is important. Owing to rapidly developing technology, the majority of the tested devices included in this review are now out of date or discontinued. The nature of the consumer technology market is such that updated product iterations are commissioned even before the original iteration of a device is released. For example, the newest Apple Watch included in the review is the Series 2 watch. The Series 5 watch was released in the fall of 2019. The results are similar for all devices and brands; the Fitbit Charge HR is a popular model for validity and reliability studies, likely because of its moderate price point (approximately US \$150) compared with more expensive models (eg, Garmin Fenix 5, approximately US \$500). Given the current device specialization, device relevance (eg, swimming or sleep-specific watches), and price difference between devices, continuing to conduct the types of reliability and validity studies reported here will be a challenge. The increasing pace of device release combined with device specialization makes this type of research challenging.

Second, few studies reported on how data were acquired from the devices. We believe this has implications for the scale of and usability of the data collected. For example, in order to collect data, we infer that some studies counted the steps recorded on the device in short time intervals instead of connecting the device to a platform after recording. Other studies exported and downloaded data from user accounts on the brand website, while others collected data from the brand API. Collecting data from the device API is the best and most scalable method for physical activity researchers when using wearable device data. In order to do so, we must develop interdisciplinary collaborations and open source tools to allow these data to be collected (eg, Open mHealth) [180].

Third, the algorithms used in consumer wearables are constantly changing based on sensor development and technological advances. Companies can update their devices' firmware and algorithm at any time. When the device is synced, the firmware is updated. Feehan et al discussed the importance of firmware updates in their review [13]. While we believe this is important, it is clear that companies must be more open about the algorithms they are using to estimate steps, heart rate, and energy expenditure. Given the continuing release of new devices, firmware and algorithm updates to existing devices,

and lack of availability of raw data, we believe researchers may need to shift focus from traditional reliability and validity research to studies that can provide open estimates for physical activity intensities or sleep standardized across devices. These studies will need to use device APIs and machine learning methods in collaboration with interdisciplinary teams in order to move the field forward.

### Limitations

Over the course of time that it took to complete this review, much has changed with market share, technology, and even research methodologies. Though the market share of companies was a large determining factor of what devices were included in this review, the consumer wearable market is volatile. On November 1, 2019, Google purchased Fitbit for US \$2.3 billion, a massive shift for the consumer wearable device market [181,182]. Further to this limitation is the ever-changing nature of consumer technology. As Table 1 shows, many of the devices utilized in the studies included in this review are so out of date that they are no longer available on the market. There is some potential for bias when including only English language studies in systematic reviews. However, studies have shown that the effect may be small in general but may be difficult to measure for an individual systematic review [183,184].

### Conclusion

This systematic review of 158 publications included assessments of consumer wearable devices from nine brands (Apple Inc, Fitbit, Garmin, Mio, Misfit, Polar, Samsung, Withings, and Xiaomi), with a focus on the reliability and validity of the devices in measuring heart rate, energy expenditure, and step count. This review examined the validity of consumer wearable devices in free-living and laboratory settings and further highlighted results of the inter- and intradevice reliability of the nine consumer wearable brands. Among the studies included, Fitbit was studied the most and Xiaomi and Mio were studied the least. Apple and Samsung had the highest validity for step count, and Apple, Fitbit, and Garmin were accurate nearly 50% of the time. No brand fell within the acceptable accuracy limits for energy expenditure. Interdevice reliabilities for steps, heart rate, and calories were all very strong. Sufficient data for intradevice reliability were only available for step count, and the results showed considerable variability. There was no specific device or brand that involved a complete assessment across all measures, and no specific brand stood out as the "gold standard" in fitness wearables. This review highlights the validity and reliability of readily available wearable devices from brands and serves to guide researchers in making decisions about including them in their research. As new devices and models enter the market, up-to-date documentation can help direct their use in the research setting.

### Acknowledgments

The authors would like to thank Kristen Romme, a librarian at Memorial University, for writing and executing the searches. They would also like to thank Dr Sherri Tran who assisted with article screening as part of her medical training. Funding for this research was provided by Dr Fuller's Canada Research Chair (#950-230773).

## Authors' Contributions

DF, EC, and JF conceptualized the paper and lead the review. EC, JF, KO, MT, BS, RB, DvH, LS, HL, and NT conducted screening. DF, EC, JF, and KC conducted methodological review and data analysis. All authors contributed to writing of the manuscript and approved the submitted version of the manuscript.

## Conflicts of Interest

Author JL was employed by Garmin Inc. during the publication process but after completion of the paper. All other authors have no conflicts to declare.

## Multimedia Appendix 1

Search strategies.

[\[PDF File \(Adobe PDF File\), 35 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) risk of bias.

[\[PDF File \(Adobe PDF File\), 131 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Substudies.

[\[PDF File \(Adobe PDF File\), 11 KB-Multimedia Appendix 3\]](#)

## Multimedia Appendix 4

Study demographics.

[\[PDF File \(Adobe PDF File\), 254 KB-Multimedia Appendix 4\]](#)

## Multimedia Appendix 5

Device models.

[\[PDF File \(Adobe PDF File\), 197 KB-Multimedia Appendix 5\]](#)

## Multimedia Appendix 6

Criterion measures.

[\[PDF File \(Adobe PDF File\), 193 KB-Multimedia Appendix 6\]](#)

## Multimedia Appendix 7

Energy expenditure mean absolute percentage error stratified by heart rate measurement on the device.

[\[PDF File \(Adobe PDF File\), 5 KB-Multimedia Appendix 7\]](#)

## References

1. World Health Organization. Global recommendations on physical activity for health. World Health Organization. URL: [https://www.who.int/dietphysicalactivity/factsheet\\_recommendations/en/](https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/) [accessed 2020-08-17]
2. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. *Am J Public Health* 2011 Apr;101(4):751-758 [FREE Full text] [doi: [10.2105/AJPH.2010.194381](https://doi.org/10.2105/AJPH.2010.194381)] [Medline: [21330590](https://pubmed.ncbi.nlm.nih.gov/21330590/)]
3. Tudor-Locke C, Craig CL, Brown WJ, Clesmes SA, De CK, Giles-Corti B, et al. How many steps/day are enough? For adults. *Int J Behav Nutr Phys Act* 2011 Jul 28;8:79 [FREE Full text] [doi: [10.1186/1479-5868-8-79](https://doi.org/10.1186/1479-5868-8-79)] [Medline: [21798015](https://pubmed.ncbi.nlm.nih.gov/21798015/)]
4. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. *Am J Health Promot* 2006;21(2):85-89. [doi: [10.4278/0890-1171-21.2.85](https://doi.org/10.4278/0890-1171-21.2.85)] [Medline: [17152246](https://pubmed.ncbi.nlm.nih.gov/17152246/)]
5. Marshall SJ, Levy SS, Tudor-Locke CE, Kolkhorst FW, Wooten KM, Ji M, et al. Translating physical activity recommendations into a pedometer-based step goal: 3000 steps in 30 minutes. *Am J Prev Med* 2009 May;36(5):410-415. [doi: [10.1016/j.amepre.2009.01.021](https://doi.org/10.1016/j.amepre.2009.01.021)] [Medline: [19362695](https://pubmed.ncbi.nlm.nih.gov/19362695/)]
6. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA* 2015 Feb 10;313(6):625-626. [doi: [10.1001/jama.2014.17841](https://doi.org/10.1001/jama.2014.17841)] [Medline: [25668268](https://pubmed.ncbi.nlm.nih.gov/25668268/)]



7. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. *Health Inf Sci Syst* 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: [10.1186/2047-2501-3-S1-S1](https://doi.org/10.1186/2047-2501-3-S1-S1)] [Medline: [26019809](https://pubmed.ncbi.nlm.nih.gov/26019809/)]
8. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. *Critical Public Health* 2013 Dec;23(4):393-403. [doi: [10.1080/09581596.2013.794931](https://doi.org/10.1080/09581596.2013.794931)]
9. Costello K. Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019. Gartner, Inc. 2018. URL: <https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow> [accessed 2020-08-17]
10. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. *BMJ Open* 2016 Jul 07;6(7):e011243 [FREE Full text] [doi: [10.1136/bmjopen-2016-011243](https://doi.org/10.1136/bmjopen-2016-011243)] [Medline: [27388359](https://pubmed.ncbi.nlm.nih.gov/27388359/)]
11. Macridis S, Johnston N, Johnson S, Vallance JK. Consumer physical activity tracking device ownership and use among a population-based sample of adults. *PLoS One* 2018;13(1):e0189298 [FREE Full text] [doi: [10.1371/journal.pone.0189298](https://doi.org/10.1371/journal.pone.0189298)] [Medline: [29293532](https://pubmed.ncbi.nlm.nih.gov/29293532/)]
12. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *Int J Exerc Sci* 2018;11(7):503-515 [FREE Full text] [Medline: [29541338](https://pubmed.ncbi.nlm.nih.gov/29541338/)]
13. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. *JMIR Mhealth Uhealth* 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: [10.2196/10527](https://doi.org/10.2196/10527)] [Medline: [30093371](https://pubmed.ncbi.nlm.nih.gov/30093371/)]
14. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015 Jan 01;4:1 [FREE Full text] [doi: [10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)] [Medline: [25554246](https://pubmed.ncbi.nlm.nih.gov/25554246/)]
15. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016 Jul;75:40-46 [FREE Full text] [doi: [10.1016/j.jclinepi.2016.01.021](https://doi.org/10.1016/j.jclinepi.2016.01.021)] [Medline: [27005575](https://pubmed.ncbi.nlm.nih.gov/27005575/)]
16. Higgins PA, Straub AJ. Understanding the error of our ways: mapping the concepts of validity and reliability. *Nurs Outlook* 2006;54(1):23-29. [doi: [10.1016/j.outlook.2004.12.004](https://doi.org/10.1016/j.outlook.2004.12.004)] [Medline: [16487776](https://pubmed.ncbi.nlm.nih.gov/16487776/)]
17. Swinscow T, Campbell M. *Statistics at Square One*. Oxford: BMJ Books; 2009.
18. Replication Data for: Systematic Review of the Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate. Harvard Dataverse. URL: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/O7GQIM> [accessed 2020-08-06]
19. wearable\_systematic\_review. GitHub. URL: [https://github.com/walkabillylab/wearable\\_systematic\\_review](https://github.com/walkabillylab/wearable_systematic_review) [accessed 2020-08-06]
20. Bai Y, Hibbing P, Mantis C, Welk G. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. *J Sports Sci* 2018 Aug;36(15):1734-1741. [doi: [10.1080/02640414.2017.1412235](https://doi.org/10.1080/02640414.2017.1412235)] [Medline: [29210326](https://pubmed.ncbi.nlm.nih.gov/29210326/)]
21. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR Mhealth Uhealth* 2017 Mar 16;5(3):e34 [FREE Full text] [doi: [10.2196/mhealth.7043](https://doi.org/10.2196/mhealth.7043)] [Medline: [28302596](https://pubmed.ncbi.nlm.nih.gov/28302596/)]
22. Koshy AN, Sajeew JK, Nerlekar N, Brown AJ, Rajakariar K, Zureik M, et al. Smart watches for heart rate assessment in atrial arrhythmias. *Int J Cardiol* 2018 Sep 01;266:124-127. [doi: [10.1016/j.ijcard.2018.02.073](https://doi.org/10.1016/j.ijcard.2018.02.073)] [Medline: [29887428](https://pubmed.ncbi.nlm.nih.gov/29887428/)]
23. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *J Sports Sci* 2019 Jun;37(12):1411-1419. [doi: [10.1080/02640414.2018.1560644](https://doi.org/10.1080/02640414.2018.1560644)] [Medline: [30657025](https://pubmed.ncbi.nlm.nih.gov/30657025/)]
24. Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch™ for measuring maximal heart rate. *J Sports Sci* 2018 Jul;36(13):1447-1452. [doi: [10.1080/02640414.2017.1397282](https://doi.org/10.1080/02640414.2017.1397282)] [Medline: [29090987](https://pubmed.ncbi.nlm.nih.gov/29090987/)]
25. Bunn J, Wells E, Manor J, Webster M. Evaluation of Earbud and Wristwatch Heart Rate Monitors during Aerobic and Resistance Training. *Int J Exerc Sci* 2019;12(4):374-384 [FREE Full text] [Medline: [30899350](https://pubmed.ncbi.nlm.nih.gov/30899350/)]
26. Khushhal A, Nichols S, Evans W, Gleadall-Siddall D, Page R, O'Doherty A, et al. Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise. *Sports Med Int Open* 2017 Oct;1(6):E206-E211 [FREE Full text] [doi: [10.1055/s-0043-120195](https://doi.org/10.1055/s-0043-120195)] [Medline: [30539109](https://pubmed.ncbi.nlm.nih.gov/30539109/)]
27. Pope ZC, Lee JE, Zeng N, Gao Z. Validation of Four Smartwatches in Energy Expenditure and Heart Rate Assessment During Exergaming. *Games Health J* 2019 Jun;8(3):205-212. [doi: [10.1089/g4h.2018.0087](https://doi.org/10.1089/g4h.2018.0087)] [Medline: [31045446](https://pubmed.ncbi.nlm.nih.gov/31045446/)]
28. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot Study Assessing the Influence of Skin Type on the Heart Rate Measurements Obtained by Photoplethysmography with the Apple Watch. *J Med Syst* 2019 May 22;43(7):195. [doi: [10.1007/s10916-019-1325-2](https://doi.org/10.1007/s10916-019-1325-2)] [Medline: [31119387](https://pubmed.ncbi.nlm.nih.gov/31119387/)]
29. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS One* 2017;12(2):e0171720 [FREE Full text] [doi: [10.1371/journal.pone.0171720](https://doi.org/10.1371/journal.pone.0171720)] [Medline: [28234979](https://pubmed.ncbi.nlm.nih.gov/28234979/)]

30. Nuss KJ, Thomson EA, Courtney JB, Comstock A, Reinwald S, Blake S, et al. Assessment of Accuracy of Overall Energy Expenditure Measurements for the Fitbit Charge HR 2 and Apple Watch. *Am J Health Behav* 2019 May 01;43(3):498-505. [doi: [10.5993/ajhb.43.3.5](https://doi.org/10.5993/ajhb.43.3.5)]
31. Zhang P, Godin SD, Owens MV. Measuring the validity and reliability of the Apple Watch as a physical activity monitor. *J Sports Med Phys Fitness* 2019 May;59(5). [doi: [10.23736/s0022-4707.18.08339-1](https://doi.org/10.23736/s0022-4707.18.08339-1)]
32. Fokkema T, Kooiman T, Krijnen W, Van DS, De GM. Reliability and validity of ten consumer activity trackers depend on walking speed. *Medicine and Science in Sport and Exercise* 2017;49(4):793-800. [doi: [10.1249/mss.0000000000001146](https://doi.org/10.1249/mss.0000000000001146)]
33. Breteler MJ, Janssen JH, Spiering W, Kalkman CJ, van Solinge WW, Dohmen DA. Measuring Free-Living Physical Activity With Three Commercially Available Activity Monitors for Telemonitoring Purposes: Validation Study. *JMIR Form Res* 2019 Apr 24;3(2):e11489 [FREE Full text] [doi: [10.2196/11489](https://doi.org/10.2196/11489)] [Medline: [31017587](https://pubmed.ncbi.nlm.nih.gov/31017587/)]
34. Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, et al. Determining the Validity and Accuracy of Multiple Activity-Tracking Devices in Controlled and Free-Walking Conditions. *Am J Health Promot* 2018 Nov;32(8):1671-1678. [doi: [10.1177/0890117118763273](https://doi.org/10.1177/0890117118763273)] [Medline: [29558811](https://pubmed.ncbi.nlm.nih.gov/29558811/)]
35. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile Device Accuracy for Step Counting Across Age Groups. *JMIR Mhealth Uhealth* 2017 Jun 28;5(6):e88 [FREE Full text] [doi: [10.2196/mhealth.7870](https://doi.org/10.2196/mhealth.7870)] [Medline: [28659255](https://pubmed.ncbi.nlm.nih.gov/28659255/)]
36. Sears T, Alvalos E, Lawson S, McAlister I, Bunn J. Wrist-Worn Physical Activity Trackers Tend To Underestimate Steps During Walking. *International Journal of Exercise Science* 2017;10(5):764-773 [FREE Full text]
37. Veerabhadrapa P, Moran MD, Renninger MD, Rhudy MB, Dreisbach SB, Gift KM. Tracking Steps on Apple Watch at Different Walking Speeds. *J Gen Intern Med* 2018 Jun;33(6):795-796 [FREE Full text] [doi: [10.1007/s11606-018-4332-y](https://doi.org/10.1007/s11606-018-4332-y)] [Medline: [29633143](https://pubmed.ncbi.nlm.nih.gov/29633143/)]
38. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of Heart Rate Watches: Implications for Weight Management. *PLoS One* 2016;11(5):e0154420 [FREE Full text] [doi: [10.1371/journal.pone.0154420](https://doi.org/10.1371/journal.pone.0154420)] [Medline: [27232714](https://pubmed.ncbi.nlm.nih.gov/27232714/)]
39. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch Measurements for Heart Rate and Energy Expenditure in Patients With Cardiovascular Disease: Cross-Sectional Study. *JMIR Mhealth Uhealth* 2019 Mar 19;7(3):e11889 [FREE Full text] [doi: [10.2196/11889](https://doi.org/10.2196/11889)] [Medline: [30888332](https://pubmed.ncbi.nlm.nih.gov/30888332/)]
40. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov A, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Med Sci Sports Exerc* 2017 Aug;49(8):1697-1703. [doi: [10.1249/MSS.0000000000001284](https://doi.org/10.1249/MSS.0000000000001284)] [Medline: [28709155](https://pubmed.ncbi.nlm.nih.gov/28709155/)]
41. Lee M, Lee H, Park S. Accuracy of swimming wearable watches for estimating energy expenditure. *International Journal of Applied Sport Science* 2018;30(1):80-90. [doi: [10.24985/ijass.2018.30.1.80](https://doi.org/10.24985/ijass.2018.30.1.80)]
42. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *JMIR Mhealth Uhealth* 2018 Apr 12;6(4):e94 [FREE Full text] [doi: [10.2196/mhealth.9754](https://doi.org/10.2196/mhealth.9754)] [Medline: [29650506](https://pubmed.ncbi.nlm.nih.gov/29650506/)]
43. Boudreaux B, Hebert E, Hollander D, Williams B, Cormier C, Naquin M, et al. Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. *Med Sci Sports Exerc* 2018 Mar;50(3):624-633. [doi: [10.1249/MSS.0000000000001471](https://doi.org/10.1249/MSS.0000000000001471)] [Medline: [29189666](https://pubmed.ncbi.nlm.nih.gov/29189666/)]
44. Hwang J, Kim J, Choi K, Cho MS, Nam G, Kim Y. Assessing Accuracy of Wrist-Worn Wearable Devices in Measurement of Paroxysmal Supraventricular Tachycardia Heart Rate. *Korean Circ J* 2019 May;49(5):437-445 [FREE Full text] [doi: [10.4070/kcj.2018.0323](https://doi.org/10.4070/kcj.2018.0323)] [Medline: [30808083](https://pubmed.ncbi.nlm.nih.gov/30808083/)]
45. 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(1):e000254 [FREE Full text] [doi: [10.1136/bmjsem-2017-000254](https://doi.org/10.1136/bmjsem-2017-000254)] [Medline: [29018543](https://pubmed.ncbi.nlm.nih.gov/29018543/)]
46. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Perera S, Kowalsky RJ, Brach JS. Accuracy and Acceptability of Commercial-Grade Physical Activity Monitors in Older Adults. *J Aging Phys Act* 2019 Apr 01;27(2):222-229 [FREE Full text] [doi: [10.1123/japa.2018-0036](https://doi.org/10.1123/japa.2018-0036)] [Medline: [30117355](https://pubmed.ncbi.nlm.nih.gov/30117355/)]
47. Husted HM, Llewellyn TL. The Accuracy of Pedometers in Measuring Walking Steps on a Treadmill in College Students. *Int J Exerc Sci* 2017;10(1):146-153 [FREE Full text] [Medline: [28479955](https://pubmed.ncbi.nlm.nih.gov/28479955/)]
48. DeShaw K, Ellingson L, Bai Y, Lansing J, Perez M, Welk G. Methods for Activity Monitor Validation Studies: An Example With the Fitbit Charge. *Journal for the Measurement of Physical Behaviour* 2018;1(3):130-135. [doi: [10.1123/jmpb.2018-0017](https://doi.org/10.1123/jmpb.2018-0017)]
49. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. Comparison of wrist-worn and hip-worn activity monitors under free living conditions. *J Med Eng Technol* 2017 Apr;41(3):200-207. [doi: [10.1080/03091902.2016.1271046](https://doi.org/10.1080/03091902.2016.1271046)] [Medline: [28078908](https://pubmed.ncbi.nlm.nih.gov/28078908/)]
50. Smith JD, Guerra G, Burkholder BG. The validity and accuracy of wrist-worn activity monitors in lower-limb prosthesis users. *Disabil Rehabil* 2019 Apr 12;1-7. [doi: [10.1080/09638288.2019.1587792](https://doi.org/10.1080/09638288.2019.1587792)] [Medline: [30978125](https://pubmed.ncbi.nlm.nih.gov/30978125/)]
51. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. *Medicine & Science in Sports & Exercise* 2018;50(6):1315-1322. [doi: [10.1249/mss.0000000000001569](https://doi.org/10.1249/mss.0000000000001569)]

52. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of Different Activity Monitors to Count Steps in an Inpatient Rehabilitation Setting. *Phys Ther* 2017 May 01;97(5):581-588. [doi: [10.1093/ptj/pzx010](https://doi.org/10.1093/ptj/pzx010)] [Medline: [28339904](https://pubmed.ncbi.nlm.nih.gov/28339904/)]
53. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions. *Front Physiol* 2017;8:725 [FREE Full text] [doi: [10.3389/fphys.2017.00725](https://doi.org/10.3389/fphys.2017.00725)] [Medline: [29018355](https://pubmed.ncbi.nlm.nih.gov/29018355/)]
54. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. *Ann Intern Med* 2017 Apr 18;166(8):610-612 [FREE Full text] [doi: [10.7326/L16-0353](https://doi.org/10.7326/L16-0353)] [Medline: [28395305](https://pubmed.ncbi.nlm.nih.gov/28395305/)]
55. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *J Med Eng Technol* 2016 Aug;40(6):325-329. [doi: [10.1080/03091902.2016.1197978](https://doi.org/10.1080/03091902.2016.1197978)] [Medline: [27280592](https://pubmed.ncbi.nlm.nih.gov/27280592/)]
56. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The Accuracy of Smart Devices for Measuring Physical Activity in Daily Life: Validation Study. *JMIR Mhealth Uhealth* 2018 Dec 13;6(12):e10972 [FREE Full text] [doi: [10.2196/10972](https://doi.org/10.2196/10972)] [Medline: [30545810](https://pubmed.ncbi.nlm.nih.gov/30545810/)]
57. Poojary J, Arora E, Britto A, Polen Z, Arena R, Babu AS. Validity of Mobile-Based Technology vs Direct Observation in Measuring Number of Steps and Distance Walked in 6 Minutes. *Mayo Clin Proc* 2018 Dec;93(12):1873-1874. [doi: [10.1016/j.mayocp.2018.09.003](https://doi.org/10.1016/j.mayocp.2018.09.003)] [Medline: [30522597](https://pubmed.ncbi.nlm.nih.gov/30522597/)]
58. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. *PLoS One* 2019;14(5):e0216891 [FREE Full text] [doi: [10.1371/journal.pone.0216891](https://doi.org/10.1371/journal.pone.0216891)] [Medline: [31112585](https://pubmed.ncbi.nlm.nih.gov/31112585/)]
59. 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(1):e0211231 [FREE Full text] [doi: [10.1371/journal.pone.0211231](https://doi.org/10.1371/journal.pone.0211231)] [Medline: [30699159](https://pubmed.ncbi.nlm.nih.gov/30699159/)]
60. Keating XD, Liu J, Liu X, Shanguan R, Guan J, Chen L. Validity of Fitbit Charge 2 in Controlled College Physical Education Settings. *ICHPER-SD Journal of Research* 2012;9(2):28-35 [FREE Full text]
61. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One* 2018;13(2):e0192691 [FREE Full text] [doi: [10.1371/journal.pone.0192691](https://doi.org/10.1371/journal.pone.0192691)] [Medline: [29489850](https://pubmed.ncbi.nlm.nih.gov/29489850/)]
62. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. *Chronobiol Int* 2019 Jul;36(7):927-933. [doi: [10.1080/07420528.2019.1596947](https://doi.org/10.1080/07420528.2019.1596947)] [Medline: [30990098](https://pubmed.ncbi.nlm.nih.gov/30990098/)]
63. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study. *JMIR Mhealth Uhealth* 2018 Dec 10;6(12):e10338 [FREE Full text] [doi: [10.2196/10338](https://doi.org/10.2196/10338)] [Medline: [30530451](https://pubmed.ncbi.nlm.nih.gov/30530451/)]
64. Dondzila C, Lewis C, Lopez J, Parker T. Congruent Accuracy of Wrist-worn Activity Trackers during Controlled and Free-living Conditions. *International Journal of Exercise Science* 2018;11(7):575-584 [FREE Full text]
65. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. *Gait Posture* 2018 Jun;63:104-108. [doi: [10.1016/j.gaitpost.2018.04.034](https://doi.org/10.1016/j.gaitpost.2018.04.034)] [Medline: [29729611](https://pubmed.ncbi.nlm.nih.gov/29729611/)]
66. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative Accuracy of a Wrist-Worn Activity Tracker and a Smart Shirt for Physical Activity Assessment. *Measurement in Physical Education and Exercise Science* 2017 Jun 08;21(4):201-211. [doi: [10.1080/1091367X.2017.1331166](https://doi.org/10.1080/1091367X.2017.1331166)]
67. Jo E, Lewis K, Directo D, Kim M, Dolezal B. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. *J Sports Sci Med* 2016 Sep;15(3):540-547 [FREE Full text] [Medline: [27803634](https://pubmed.ncbi.nlm.nih.gov/27803634/)]
68. Kroll RR, Boyd JG, Maslove DM. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. *J Med Internet Res* 2016 Sep 20;18(9):e253 [FREE Full text] [doi: [10.2196/jmir.6025](https://doi.org/10.2196/jmir.6025)] [Medline: [27651304](https://pubmed.ncbi.nlm.nih.gov/27651304/)]
69. Pelizzo G, Guddo A, Puglisi A, De Silvestri A, Comparato C, Valenza M, et al. Accuracy of a Wrist-Worn Heart Rate Sensing Device during Elective Pediatric Surgical Procedures. *Children (Basel)* 2018 Mar 08;5(3) [FREE Full text] [doi: [10.3390/children5030038](https://doi.org/10.3390/children5030038)] [Medline: [29518020](https://pubmed.ncbi.nlm.nih.gov/29518020/)]
70. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation Of A Self-monitoring Tool For Use In Post-concussion Syndrome Therapy. *Medicine & Science in Sports & Exercise* 2016;48:798. [doi: [10.1249/01.mss.0000487390.22076.d1](https://doi.org/10.1249/01.mss.0000487390.22076.d1)]
71. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc Med* 2016;2(1):e000106 [FREE Full text] [doi: [10.1136/bmjsem-2015-000106](https://doi.org/10.1136/bmjsem-2015-000106)] [Medline: [27900173](https://pubmed.ncbi.nlm.nih.gov/27900173/)]
72. Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. *BMC Geriatr* 2018 May 03;18(1):103 [FREE Full text] [doi: [10.1186/s12877-018-0793-4](https://doi.org/10.1186/s12877-018-0793-4)] [Medline: [29724191](https://pubmed.ncbi.nlm.nih.gov/29724191/)]
73. Morris CE, Wessel PA, Tinius RA, Schafer MA, Maples JM. Validity of Activity Trackers in Estimating Energy Expenditure During High-Intensity Functional Training. *Res Q Exerc Sport* 2019 Sep;90(3):377-384. [doi: [10.1080/02701367.2019.1603989](https://doi.org/10.1080/02701367.2019.1603989)] [Medline: [31046630](https://pubmed.ncbi.nlm.nih.gov/31046630/)]



74. Farina N, Lowry RG. The Validity of Consumer-Level Activity Monitors in Healthy Older Adults in Free-Living Conditions. *J Aging Phys Act* 2018 Jan 01;26(1):128-135. [doi: [10.1123/japa.2016-0344](https://doi.org/10.1123/japa.2016-0344)] [Medline: [28595019](https://pubmed.ncbi.nlm.nih.gov/28595019/)]
75. Rozanski GM, Aquil A, Sivakumaran S, Mansfield A. Consumer Wearable Devices for Activity Monitoring Among Individuals After a Stroke: A Prospective Comparison. *JMIR Cardio* 2018 Jan 04;2(1):e1 [FREE Full text] [doi: [10.2196/cardio.8199](https://doi.org/10.2196/cardio.8199)] [Medline: [31758760](https://pubmed.ncbi.nlm.nih.gov/31758760/)]
76. Voss C, Gardner RF, Dean PH, Harris KC. Validity of Commercial Activity Trackers in Children With Congenital Heart Disease. *Can J Cardiol* 2017 Jun;33(6):799-805. [doi: [10.1016/j.cjca.2016.11.024](https://doi.org/10.1016/j.cjca.2016.11.024)] [Medline: [28347581](https://pubmed.ncbi.nlm.nih.gov/28347581/)]
77. Brazendale K, Decker L, Hunt ET, Perry MW, Brazendale AB, Weaver RG, et al. Validity and Wearability of Consumer-based Fitness Trackers in Free-living Children. *Int J Exerc Sci* 2019;12(5):471-482 [FREE Full text] [Medline: [30899354](https://pubmed.ncbi.nlm.nih.gov/30899354/)]
78. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions. *JMIR Mhealth Uhealth* 2017 Oct 20;5(10):e157 [FREE Full text] [doi: [10.2196/mhealth.8233](https://doi.org/10.2196/mhealth.8233)] [Medline: [29055881](https://pubmed.ncbi.nlm.nih.gov/29055881/)]
79. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent Validity of Wearable Activity Trackers Under Free-Living Conditions. *Journal of Strength and Conditioning Research* 2017;31(4):1097-1106. [doi: [10.1519/jsc.0000000000001571](https://doi.org/10.1519/jsc.0000000000001571)]
80. Chow J, Thom J, Wewege M, Ward R, Parmenter B. Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. *Gait Posture* 2017 Sep;57:199-203. [doi: [10.1016/j.gaitpost.2017.06.012](https://doi.org/10.1016/j.gaitpost.2017.06.012)] [Medline: [28666177](https://pubmed.ncbi.nlm.nih.gov/28666177/)]
81. Tam KM, Cheung SY. Validation of Electronic Activity Monitor Devices During Treadmill Walking. *Telemed J E Health* 2018 Oct;24(10):782-789. [doi: [10.1089/tmj.2017.0263](https://doi.org/10.1089/tmj.2017.0263)] [Medline: [29364065](https://pubmed.ncbi.nlm.nih.gov/29364065/)]
82. Tophøj KH, Petersen MG, Sæbye C, Baad-Hansen T, Wagner S. Validity and Reliability Evaluation of Four Commercial Activity Trackers' Step Counting Performance. *Telemed J E Health* 2018 Sep;24(9):669-677. [doi: [10.1089/tmj.2017.0264](https://doi.org/10.1089/tmj.2017.0264)] [Medline: [29303680](https://pubmed.ncbi.nlm.nih.gov/29303680/)]
83. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. *Med Eng Phys* 2014 Jun;36(6):659-669 [FREE Full text] [doi: [10.1016/j.medengphy.2014.02.006](https://doi.org/10.1016/j.medengphy.2014.02.006)] [Medline: [24656871](https://pubmed.ncbi.nlm.nih.gov/24656871/)]
84. Phillips LJ, Petroski GF, Markis NE. A Comparison of Accelerometer Accuracy in Older Adults. *Res Gerontol Nurs* 2015;8(5):213-219. [doi: [10.3928/19404921-20150429-03](https://doi.org/10.3928/19404921-20150429-03)] [Medline: [25942386](https://pubmed.ncbi.nlm.nih.gov/25942386/)]
85. Adam Noah J, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *J Med Eng Technol* 2013 Oct;37(7):456-462. [doi: [10.3109/03091902.2013.831135](https://doi.org/10.3109/03091902.2013.831135)] [Medline: [24007317](https://pubmed.ncbi.nlm.nih.gov/24007317/)]
86. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc* 2013 Nov;45(11):2105-2112 [FREE Full text] [doi: [10.1249/MSS.0b013e318299d2eb](https://doi.org/10.1249/MSS.0b013e318299d2eb)] [Medline: [23669877](https://pubmed.ncbi.nlm.nih.gov/23669877/)]
87. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Kozey Keadle S, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. *J Phys Act Health* 2015 Feb;12(2):149-154. [doi: [10.1123/jpah.2012-0495](https://doi.org/10.1123/jpah.2012-0495)] [Medline: [24770438](https://pubmed.ncbi.nlm.nih.gov/24770438/)]
88. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. *JMIR Mhealth Uhealth* 2017 Aug 11;5(8):e106. [doi: [10.2196/mhealth.6321](https://doi.org/10.2196/mhealth.6321)]
89. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? *Eur J Sport Sci* 2017 Apr;17(3):360-368. [doi: [10.1080/17461391.2016.1255261](https://doi.org/10.1080/17461391.2016.1255261)] [Medline: [27912681](https://pubmed.ncbi.nlm.nih.gov/27912681/)]
90. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities. *J Gerontol A Biol Sci Med Sci* 2017 Feb;72(2):229-236 [FREE Full text] [doi: [10.1093/gerona/glw098](https://doi.org/10.1093/gerona/glw098)] [Medline: [27257217](https://pubmed.ncbi.nlm.nih.gov/27257217/)]
91. Gilmore SJ, Davidson M, Hahne AJ, McClelland JA. The validity of using activity monitors to detect step count after lumbar fusion surgery. *Disabil Rehabil* 2020 Mar;42(6):863-868. [doi: [10.1080/09638288.2018.1509140](https://doi.org/10.1080/09638288.2018.1509140)] [Medline: [30326754](https://pubmed.ncbi.nlm.nih.gov/30326754/)]
92. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait Posture* 2016 Jul;48:36-41. [doi: [10.1016/j.gaitpost.2016.04.025](https://doi.org/10.1016/j.gaitpost.2016.04.025)] [Medline: [27477705](https://pubmed.ncbi.nlm.nih.gov/27477705/)]
93. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. *Br J Sports Med* 2018 Jul;52(13):844-850. [doi: [10.1136/bjsports-2016-096990](https://doi.org/10.1136/bjsports-2016-096990)] [Medline: [28483930](https://pubmed.ncbi.nlm.nih.gov/28483930/)]
94. Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and Reliability of the Fitbit Flex™ and Actigraph Gt3X+ At Jogging and Running Speeds. *Intl J Sports Phys Ther* 2018 Aug;13(5):860-870. [doi: [10.26603/ijst20180860](https://doi.org/10.26603/ijst20180860)]
95. Kendall B, Bellovary B, Gothe NP. Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. *J Sports Sci* 2019 Jan;37(1):42-49. [doi: [10.1080/02640414.2018.1481723](https://doi.org/10.1080/02640414.2018.1481723)] [Medline: [29863968](https://pubmed.ncbi.nlm.nih.gov/29863968/)]

96. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil* 2015;7:24 [FREE Full text] [doi: [10.1186/s13102-015-0018-5](https://doi.org/10.1186/s13102-015-0018-5)] [Medline: [26464801](https://pubmed.ncbi.nlm.nih.gov/26464801/)]
97. Montes J, Young J, Tandy R, Navalta J. Fitbit Flex Energy Expenditure and Step Count Evaluation. *Journal Exercise Physiology Online* 2017;20(5):134 [FREE Full text]
98. Sala DA, Grissom HE, Delsole EM, Chu ML, Godfried DH, Bhattacharyya S, et al. Measuring ambulation with wrist-based and hip-based activity trackers for children with cerebral palsy. *Dev Med Child Neurol* 2019 Nov;61(11):1309-1313. [doi: [10.1111/dmcn.14207](https://doi.org/10.1111/dmcn.14207)] [Medline: [30883727](https://pubmed.ncbi.nlm.nih.gov/30883727/)]
99. Schmal H, Holsgaard-Larsen A, Izadpanah K, Brønd JC, Madsen CF, Lauritsen J. Validation of Activity Tracking Procedures in Elderly Patients after Operative Treatment of Proximal Femur Fractures. *Rehabil Res Pract* 2018;2018:3521271 [FREE Full text] [doi: [10.1155/2018/3521271](https://doi.org/10.1155/2018/3521271)] [Medline: [30018824](https://pubmed.ncbi.nlm.nih.gov/30018824/)]
100. Balto JM, Kinnett-Hopkins DL, Motl RW. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. *Mult Scler J Exp Transl Clin* 2016;2:2055217316634754 [FREE Full text] [doi: [10.1177/2055217316634754](https://doi.org/10.1177/2055217316634754)] [Medline: [28607720](https://pubmed.ncbi.nlm.nih.gov/28607720/)]
101. 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(9):e0161224 [FREE Full text] [doi: [10.1371/journal.pone.0161224](https://doi.org/10.1371/journal.pone.0161224)] [Medline: [27589592](https://pubmed.ncbi.nlm.nih.gov/27589592/)]
102. Ummels D, Beekman E, Theunissen K, Braun S, Beurskens AJ. 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* 2018 Apr 02;6(4):e70 [FREE Full text] [doi: [10.2196/mhealth.8524](https://doi.org/10.2196/mhealth.8524)] [Medline: [29610110](https://pubmed.ncbi.nlm.nih.gov/29610110/)]
103. Nelson M, Kaminsky L, Dickin D, Montoye A. Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. *Med Sci Sports Exerc* 2016 Aug;48(8):1619-1628. [doi: [10.1249/MSS.0000000000000933](https://doi.org/10.1249/MSS.0000000000000933)] [Medline: [27015387](https://pubmed.ncbi.nlm.nih.gov/27015387/)]
104. Murakami H, Kawakami R, Nakae 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 May 01;176(5):702-703. [doi: [10.1001/jamainternmed.2016.0152](https://doi.org/10.1001/jamainternmed.2016.0152)] [Medline: [26999758](https://pubmed.ncbi.nlm.nih.gov/26999758/)]
105. Bai Y, Welk GJ, Nam YH, Lee JA, Lee J, Kim Y, et al. Comparison of Consumer and Research Monitors under Semistructured Settings. *Medicine & Science in Sports & Exercise* 2016;48(1):151-158. [doi: [10.1249/mss.0000000000000727](https://doi.org/10.1249/mss.0000000000000727)]
106. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. *Eur J Prev Cardiol* 2016 Sep;23(14):1476-1485. [doi: [10.1177/2047487316634883](https://doi.org/10.1177/2047487316634883)] [Medline: [26907794](https://pubmed.ncbi.nlm.nih.gov/26907794/)]
107. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLoS One* 2017;12(2):e0172535 [FREE Full text] [doi: [10.1371/journal.pone.0172535](https://doi.org/10.1371/journal.pone.0172535)] [Medline: [28234953](https://pubmed.ncbi.nlm.nih.gov/28234953/)]
108. 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 Sep 19;4(3):e110 [FREE Full text] [doi: [10.2196/mhealth.6281](https://doi.org/10.2196/mhealth.6281)] [Medline: [27644334](https://pubmed.ncbi.nlm.nih.gov/27644334/)]
109. Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *J Neurol* 2017 Feb;264(2):316-326 [FREE Full text] [doi: [10.1007/s00415-016-8334-6](https://doi.org/10.1007/s00415-016-8334-6)] [Medline: [27896433](https://pubmed.ncbi.nlm.nih.gov/27896433/)]
110. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *J Sci Med Sport* 2017 Jun;20(6):578-582. [doi: [10.1016/j.jsams.2016.10.015](https://doi.org/10.1016/j.jsams.2016.10.015)] [Medline: [27887786](https://pubmed.ncbi.nlm.nih.gov/27887786/)]
111. Rowe VT, Neville M. Measuring Reliability of Movement With Accelerometry: Fitbit Versus ActiGraph. *Am J Occup Ther* 2019;73(2):7302205150p1-7302205150p6. [doi: [10.5014/ajot.2019.030692](https://doi.org/10.5014/ajot.2019.030692)] [Medline: [30915976](https://pubmed.ncbi.nlm.nih.gov/30915976/)]
112. St-Laurent A, Mony MM, Mathieu M, Ruchat SM. Validation of the Fitbit Zip and Fitbit Flex with pregnant women in free-living conditions. *J Med Eng Technol* 2018 May;42(4):259-264. [doi: [10.1080/03091902.2018.1472822](https://doi.org/10.1080/03091902.2018.1472822)] [Medline: [30198806](https://pubmed.ncbi.nlm.nih.gov/30198806/)]
113. Clevenger KA, Molesky MJ, Vusich J, Montoye AH. Free-Living Comparison of Physical Activity and Sleep Data from Fitbit Activity Trackers Worn on the Dominant and Nondominant Wrists. *Measurement in Physical Education and Exercise Science* 2019 Feb 12;23(2):194-204. [doi: [10.1080/1091367x.2019.1577737](https://doi.org/10.1080/1091367x.2019.1577737)]
114. Chen MD, Kuo C, Pellegrini C, Hsu M. Accuracy of Wristband Activity Monitors during Ambulation and Activities. *Med Sci Sports Exerc* 2016 Oct;48(10):1942-1949. [doi: [10.1249/MSS.0000000000000984](https://doi.org/10.1249/MSS.0000000000000984)] [Medline: [27183123](https://pubmed.ncbi.nlm.nih.gov/27183123/)]
115. Daligadu J, Pollock CL, Carlaw K, Chin M, Haynes A, Thevarajah Kopal T, et al. Validation of the Fitbit Flex in an Acute Post-Cardiac Surgery Patient Population. *Physiother Can* 2018;70(4):314-320 [FREE Full text] [doi: [10.3138/ptc.2017-34](https://doi.org/10.3138/ptc.2017-34)] [Medline: [30745716](https://pubmed.ncbi.nlm.nih.gov/30745716/)]

116. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol* 2015 Apr 15;185:138-140 [FREE Full text] [doi: [10.1016/j.ijcard.2015.03.038](https://doi.org/10.1016/j.ijcard.2015.03.038)] [Medline: [25795203](https://pubmed.ncbi.nlm.nih.gov/25795203/)]
117. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, et al. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. *BMC Res Notes* 2016 Apr 12;9:213 [FREE Full text] [doi: [10.1186/s13104-016-2020-8](https://doi.org/10.1186/s13104-016-2020-8)] [Medline: [27068022](https://pubmed.ncbi.nlm.nih.gov/27068022/)]
118. Battenberg AK, Donohoe S, Robertson N, Schmalzried TP. The accuracy of personal activity monitoring devices. *Seminars in Arthroplasty* 2017 Jun;28(2):71-75. [doi: [10.1053/j.sart.2017.07.006](https://doi.org/10.1053/j.sart.2017.07.006)]
119. Prieto-Centurion V, Bracken N, Norwick L, Zaidi F, Mutso AA, Morken V, et al. Can Commercially Available Pedometers Be Used For Physical Activity Monitoring In Patients With COPD Following Exacerbations? *Chronic Obstr Pulm Dis* 2016;3(3):636-642 [FREE Full text] [doi: [10.15326/jcopdf.3.3.2015.0164](https://doi.org/10.15326/jcopdf.3.3.2015.0164)] [Medline: [27547817](https://pubmed.ncbi.nlm.nih.gov/27547817/)]
120. Arch ES, Sions JM, Horne J, Bodt BA. Step count accuracy of StepWatch and FitBit One™ among individuals with a unilateral transtibial amputation. *Prosthet Orthot Int* 2018 Oct;42(5):518-526. [doi: [10.1177/0309364618767138](https://doi.org/10.1177/0309364618767138)] [Medline: [29623810](https://pubmed.ncbi.nlm.nih.gov/29623810/)]
121. Klassen T, Simpson L, Lim S, Louie D, Parappilly B, Sakakibara B, et al. "Stepping Up" Activity Poststroke: Ankle-Positioned Accelerometer Can Accurately Record Steps During Slow Walking. *Phys Ther* 2016 Mar;96(3):355-360 [FREE Full text] [doi: [10.2522/ptj.20140611](https://doi.org/10.2522/ptj.20140611)] [Medline: [26251478](https://pubmed.ncbi.nlm.nih.gov/26251478/)]
122. O'Connell S, ÓLaighin G, Kelly L, Murphy E, Beirne S, Burke N, et al. These Shoes Are Made for Walking: Sensitivity Performance Evaluation of Commercial Activity Monitors under the Expected Conditions and Circumstances Required to Achieve the International Daily Step Goal of 10,000 Steps. *PLoS One* 2016;11(5):e0154956 [FREE Full text] [doi: [10.1371/journal.pone.0154956](https://doi.org/10.1371/journal.pone.0154956)] [Medline: [27167121](https://pubmed.ncbi.nlm.nih.gov/27167121/)]
123. O'Connell S, ÓLaighin G, Quinlan LR. When a Step Is Not a Step! Specificity Analysis of Five Physical Activity Monitors. *PLoS One* 2017;12(1):e0169616 [FREE Full text] [doi: [10.1371/journal.pone.0169616](https://doi.org/10.1371/journal.pone.0169616)] [Medline: [28085918](https://pubmed.ncbi.nlm.nih.gov/28085918/)]
124. Simpson L, Eng J, Klassen T, Lim S, Louie D, Parappilly B, et al. Capturing step counts at slow walking speeds in older adults: comparison of ankle and waist placement of measuring device. *J Rehabil Med* 2015 Oct 05;47(9):830-835 [FREE Full text] [doi: [10.2340/16501977-1993](https://doi.org/10.2340/16501977-1993)] [Medline: [26181670](https://pubmed.ncbi.nlm.nih.gov/26181670/)]
125. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport* 2014 Sep;17(5):496-500. [doi: [10.1016/j.jsams.2013.10.241](https://doi.org/10.1016/j.jsams.2013.10.241)] [Medline: [24268570](https://pubmed.ncbi.nlm.nih.gov/24268570/)]
126. Klassen TD, Semrau JA, Dukelow SP, Bayley MT, Hill MD, Eng JJ. Consumer-Based Physical Activity Monitor as a Practical Way to Measure Walking Intensity During Inpatient Stroke Rehabilitation. *Stroke* 2017 Sep;48(9):2614-2617. [doi: [10.1161/strokeaha.117.018175](https://doi.org/10.1161/strokeaha.117.018175)]
127. Leth S, Hansen J, Nielsen O, Dinesen B. Evaluation of Commercial Self-Monitoring Devices for Clinical Purposes: Results from the Future Patient Trial, Phase I. *Sensors (Basel)* 2017 Jan 22;17(1) [FREE Full text] [doi: [10.3390/s17010211](https://doi.org/10.3390/s17010211)] [Medline: [28117736](https://pubmed.ncbi.nlm.nih.gov/28117736/)]
128. Storm FA, Heller BW, Mazzà C. Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS One* 2015;10(3):e0118723 [FREE Full text] [doi: [10.1371/journal.pone.0118723](https://doi.org/10.1371/journal.pone.0118723)] [Medline: [25789630](https://pubmed.ncbi.nlm.nih.gov/25789630/)]
129. Lee J, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc* 2014 Sep;46(9):1840-1848. [doi: [10.1249/MSS.0000000000000287](https://doi.org/10.1249/MSS.0000000000000287)] [Medline: [24777201](https://pubmed.ncbi.nlm.nih.gov/24777201/)]
130. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J Med Eng Technol* 2017 Apr;41(3):208-215. [doi: [10.1080/03091902.2016.1253795](https://doi.org/10.1080/03091902.2016.1253795)] [Medline: [27919170](https://pubmed.ncbi.nlm.nih.gov/27919170/)]
131. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act* 2015 Mar 27;12:42 [FREE Full text] [doi: [10.1186/s12966-015-0201-9](https://doi.org/10.1186/s12966-015-0201-9)] [Medline: [25890168](https://pubmed.ncbi.nlm.nih.gov/25890168/)]
132. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. *J Med Internet Res* 2016 Sep 07;18(9):e239 [FREE Full text] [doi: [10.2196/jmir.5531](https://doi.org/10.2196/jmir.5531)] [Medline: [27604226](https://pubmed.ncbi.nlm.nih.gov/27604226/)]
133. Hamari L, Kullberg T, Ruohonen J, Heinonen O, Díaz-Rodríguez N, Lilius J, et al. Physical activity among children: objective measurements using Fitbit One and ActiGraph. *BMC Res Notes* 2017 Apr 20;10(1):161 [FREE Full text] [doi: [10.1186/s13104-017-2476-1](https://doi.org/10.1186/s13104-017-2476-1)] [Medline: [28427441](https://pubmed.ncbi.nlm.nih.gov/28427441/)]
134. Hui J, Heyden R, Bao T, Accettone N, McBay C, Richardson J, et al. Validity of the Fitbit One for Measuring Activity in Community-Dwelling Stroke Survivors. *Physiother Can* 2018;70(1):81-89 [FREE Full text] [doi: [10.3138/ptc.2016-40.ep](https://doi.org/10.3138/ptc.2016-40.ep)] [Medline: [29434422](https://pubmed.ncbi.nlm.nih.gov/29434422/)]
135. Middelweerd A, Van Der Ploeg HP, Van Halteren A, Twisk JW, Brug J, Te Velde SJ. A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals. *Med Sci Sports Exerc* 2017 Jun;49(6):1270-1279. [doi: [10.1249/MSS.0000000000001225](https://doi.org/10.1249/MSS.0000000000001225)] [Medline: [28511193](https://pubmed.ncbi.nlm.nih.gov/28511193/)]
136. Van Blarigan EL, Kenfield SA, Tantum L, Cadmus-Bertram LA, Carroll PR, Chan JM. The Fitbit One Physical Activity Tracker in Men With Prostate Cancer: Validation Study. *JMIR Cancer* 2017 Apr 18;3(1):e5 [FREE Full text] [doi: [10.2196/cancer.6935](https://doi.org/10.2196/cancer.6935)] [Medline: [28420602](https://pubmed.ncbi.nlm.nih.gov/28420602/)]



137. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. *Med Sci Sports Exerc* 2016 Mar;48(3):457-465 [FREE Full text] [doi: [10.1249/MSS.0000000000000778](https://doi.org/10.1249/MSS.0000000000000778)] [Medline: [26484953](https://pubmed.ncbi.nlm.nih.gov/26484953/)]
138. Duclos NC, Aguiar LT, Aissaoui R, Faria CD, Nadeau S, Duclos C. Activity Monitor Placed at the Nonparetic Ankle Is Accurate in Measuring Step Counts During Community Walking in Poststroke Individuals: A Validation Study. *PM R* 2019 Sep;11(9):963-971. [doi: [10.1002/pmrj.12080](https://doi.org/10.1002/pmrj.12080)] [Medline: [30688026](https://pubmed.ncbi.nlm.nih.gov/30688026/)]
139. Bunn JA, Jones C, Oliviera A, Webster MJ. Assessment of step accuracy using the Consumer Technology Association standard. *J Sports Sci* 2019 Feb 29;37(3):244-248. [doi: [10.1080/02640414.2018.1491941](https://doi.org/10.1080/02640414.2018.1491941)] [Medline: [29958058](https://pubmed.ncbi.nlm.nih.gov/29958058/)]
140. Navalta J, Montes J, Bodell N, Aguilar C, Lujan A, Guzman G, et al. Wearable Device Validity in Determining Step Count During Hiking and Trail Running. *Journal for the Measurement of Physical Behaviour* 2018;1(2):86-93. [doi: [10.1123/jmpb.2018-0010](https://doi.org/10.1123/jmpb.2018-0010)]
141. Wendel N, Macpherson C, Webber K, Hendron K, DeAngelis T, Colon-Semenza C, et al. Accuracy of Activity Trackers in Parkinson Disease: Should We Prescribe Them? *Phys Ther* 2018 Aug 01;98(8):705-714. [doi: [10.1093/ptj/pzy054](https://doi.org/10.1093/ptj/pzy054)] [Medline: [29718452](https://pubmed.ncbi.nlm.nih.gov/29718452/)]
142. Thiebaud RS, Funk MD, Patton JC, Massey BL, Shay TE, Schmidt MG, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digit Health* 2018 Apr 13;4:2055207618770322 [FREE Full text] [doi: [10.1177/2055207618770322](https://doi.org/10.1177/2055207618770322)] [Medline: [29942628](https://pubmed.ncbi.nlm.nih.gov/29942628/)]
143. Pribyslavská V, Caputo JL, Coons JM, Barry VW. Impact of EPOC adjustment on estimation of energy expenditure using activity monitors. *J Med Eng Technol* 2018 May 18;42(4):265-273. [doi: [10.1080/03091902.2018.1472823](https://doi.org/10.1080/03091902.2018.1472823)] [Medline: [29911930](https://pubmed.ncbi.nlm.nih.gov/29911930/)]
144. Fulk G, Combs S, Danks K, Nirider C, Raja B, Reisman D. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. *Phys Ther* 2014 Feb;94(2):222-229. [doi: [10.2522/ptj.20120525](https://doi.org/10.2522/ptj.20120525)] [Medline: [24052577](https://pubmed.ncbi.nlm.nih.gov/24052577/)]
145. Park W, Lee VJ, Ku B, Tanaka H. Effect of walking speed and placement position interactions in determining the accuracy of various newer pedometers. *Journal of Exercise Science & Fitness* 2014 Jun;12(1):31-37. [doi: [10.1016/j.jesf.2014.01.003](https://doi.org/10.1016/j.jesf.2014.01.003)]
146. Stackpool C, Porcari J, Mikat R, Gillette C, Foster C. The Accuracy of Various Activity Trackers in Estimating Steps Taken and Energy Expenditure. *J Fit Res* 2014;3(3):32-48 [FREE Full text]
147. Wong CK, Mentis HM, Kuber R. The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. *Gait Posture* 2018 Jan;59:177-181. [doi: [10.1016/j.gaitpost.2017.10.010](https://doi.org/10.1016/j.gaitpost.2017.10.010)] [Medline: [29049964](https://pubmed.ncbi.nlm.nih.gov/29049964/)]
148. Gusmer R, Bosch T, Watkins A, Ostrem J, Dengel D. Comparison of FitBit® Ultra to ActiGraph™ GT1M for Assessment of Physical Activity in Young Adults During Treadmill Walking. *TOSMJ* 2014 Apr 04;8(1):11-15. [doi: [10.2174/1874387001408010011](https://doi.org/10.2174/1874387001408010011)]
149. Paul SS, Tiedemann A, Hassett LM, Ramsay E, Kirkham C, Chagpar S, et al. Validity of the activity tracker for measuring steps in community-dwelling older adults. *BMJ Open Sport Exerc Med* 2015;1(1):e000013 [FREE Full text] [doi: [10.1136/bmjsem-2015-000013](https://doi.org/10.1136/bmjsem-2015-000013)] [Medline: [27900119](https://pubmed.ncbi.nlm.nih.gov/27900119/)]
150. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. *Physiother Theory Pract* 2017 Oct 04;33(10):788-796. [doi: [10.1080/09593985.2017.1354412](https://doi.org/10.1080/09593985.2017.1354412)] [Medline: [28777710](https://pubmed.ncbi.nlm.nih.gov/28777710/)]
151. Sharp CA, Mackintosh KA, Erjavec M, Pascoe DM, Horne PJ. Validity and reliability of the Fitbit Zip as a measure of preschool children's step count. *BMJ Open Sport Exerc Med* 2017;3(1):e000272 [FREE Full text] [doi: [10.1136/bmjsem-2017-000272](https://doi.org/10.1136/bmjsem-2017-000272)] [Medline: [29081984](https://pubmed.ncbi.nlm.nih.gov/29081984/)]
152. Singh AK, Farmer C, Van Den Berg ML, Killington M, Barr CJ. Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. *Disabil Health J* 2016 Apr;9(2):320-323. [doi: [10.1016/j.dhjo.2015.10.011](https://doi.org/10.1016/j.dhjo.2015.10.011)] [Medline: [26905972](https://pubmed.ncbi.nlm.nih.gov/26905972/)]
153. Appelboom G, Taylor BE, Bruce E, Bassile CC, Malakidis C, Yang A, et al. Mobile Phone-Connected Wearable Motion Sensors to Assess Postoperative Mobilization. *JMIR Mhealth Uhealth* 2015 Jul 28;3(3):e78 [FREE Full text] [doi: [10.2196/mhealth.3785](https://doi.org/10.2196/mhealth.3785)] [Medline: [26220691](https://pubmed.ncbi.nlm.nih.gov/26220691/)]
154. Thorup CB, Andreasen JJ, Sørensen EE, GrønkJær M, Dinesen BI, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. *BMJ Open* 2017 Mar 31;7(3):e011742 [FREE Full text] [doi: [10.1136/bmjopen-2016-011742](https://doi.org/10.1136/bmjopen-2016-011742)] [Medline: [28363918](https://pubmed.ncbi.nlm.nih.gov/28363918/)]
155. Mooses K, Oja M, Reisberg S, Vilo J, Kull M. Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study. *BMC Public Health* 2018 Jul 11;18(1):858 [FREE Full text] [doi: [10.1186/s12889-018-5752-7](https://doi.org/10.1186/s12889-018-5752-7)] [Medline: [29996797](https://pubmed.ncbi.nlm.nih.gov/29996797/)]
156. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fiband Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Res Notes* 2014 Dec 23;7:952 [FREE Full text] [doi: [10.1186/1756-0500-7-952](https://doi.org/10.1186/1756-0500-7-952)] [Medline: [25539733](https://pubmed.ncbi.nlm.nih.gov/25539733/)]
157. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Res Notes* 2016 Sep 21;9(1):448 [FREE Full text] [doi: [10.1186/s13104-016-2253-6](https://doi.org/10.1186/s13104-016-2253-6)] [Medline: [27655477](https://pubmed.ncbi.nlm.nih.gov/27655477/)]
158. Beevi FH, Miranda J, Pedersen CF, Wagner S. An Evaluation of Commercial Pedometers for Monitoring Slow Walking Speed Populations. *Telemed J E Health* 2016 May;22(5):441-449. [doi: [10.1089/tmj.2015.0120](https://doi.org/10.1089/tmj.2015.0120)] [Medline: [26451900](https://pubmed.ncbi.nlm.nih.gov/26451900/)]

159. Boolani A, Towler C, LeCours B, Blank H, Larue J, Fulk G. Accuracy of 6 Commercially Available Activity Monitors in Measuring Heart Rate, Caloric Expenditure, Steps Walked, and Distance Traveled. *Cardiopulmonary Physical Therapy Journal* 2019;30(4):153-161. [doi: [10.1097/cpt.000000000000107](https://doi.org/10.1097/cpt.000000000000107)]
160. Chandrasekar A, Hensor EM, Mackie SL, Backhouse MR, Harris E. Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica. *Gait Posture* 2018 Mar;61:339-345. [doi: [10.1016/j.gaitpost.2018.01.035](https://doi.org/10.1016/j.gaitpost.2018.01.035)] [Medline: [29427859](https://pubmed.ncbi.nlm.nih.gov/29427859/)]
161. Haegele J, Brian A, Wolf D. Accuracy of the Fitbit Zip for Measuring Steps for Adolescents With Visual Impairments. *Adapt Phys Activ Q* 2017 Apr;34(2):195-200. [doi: [10.1123/apaq.2016-0055](https://doi.org/10.1123/apaq.2016-0055)] [Medline: [28556761](https://pubmed.ncbi.nlm.nih.gov/28556761/)]
162. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. *J Med Eng Technol* 2017 Aug 04;41(6):480-485. [doi: [10.1080/03091902.2017.1333166](https://doi.org/10.1080/03091902.2017.1333166)] [Medline: [28675070](https://pubmed.ncbi.nlm.nih.gov/28675070/)]
163. Støve MP, Haucke E, Nymann ML, Sigurdsson T, Larsen BT. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. *J Sports Sci* 2019 Apr 17;37(8):895-901. [doi: [10.1080/02640414.2018.1535563](https://doi.org/10.1080/02640414.2018.1535563)] [Medline: [30326780](https://pubmed.ncbi.nlm.nih.gov/30326780/)]
164. Mahendran N, Kuys SS, Downie E, Ng P, Brauer SG. Are Accelerometers and GPS Devices Valid, Reliable and Feasible Tools for Measurement of Community Ambulation After Stroke? *Brain Impairment* 2016 Jun 08;17(2):151-161. [doi: [10.1017/brimp.2016.13](https://doi.org/10.1017/brimp.2016.13)]
165. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. *J Med Eng Technol* 2016 Aug;40(6):298-306. [doi: [10.1080/03091902.2016.1193238](https://doi.org/10.1080/03091902.2016.1193238)] [Medline: [27266422](https://pubmed.ncbi.nlm.nih.gov/27266422/)]
166. Woodman J, Crouter S, Bassett D, Fitzhugh E, Boyer W. Accuracy of Consumer Monitors for Estimating Energy Expenditure and Activity Type. *Med Sci Sports Exerc* 2017 Feb;49(2):371-377. [doi: [10.1249/MSS.0000000000001090](https://doi.org/10.1249/MSS.0000000000001090)] [Medline: [27580155](https://pubmed.ncbi.nlm.nih.gov/27580155/)]
167. Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F. Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. *Acta Gymnica* 2016 Sep 30;46(3):129-135. [doi: [10.5507/ag.2016.014](https://doi.org/10.5507/ag.2016.014)]
168. Šimůnek A, Dygrýn J, Jakubec L, Neuls F, Frömel K, Welk G. Validity of Garmin Vivofit 1 and Garmin Vivofit 3 for School-Based Physical Activity Monitoring. *Pediatr Exerc Sci* 2019 Feb 01;31(1):130-136. [doi: [10.1123/pes.2018-0019](https://doi.org/10.1123/pes.2018-0019)] [Medline: [30304974](https://pubmed.ncbi.nlm.nih.gov/30304974/)]
169. Ehrler F, Weber C, Lovis C. Influence of Pedometer Position on Pedometer Accuracy at Various Walking Speeds: A Comparative Study. *J Med Internet Res* 2016 Oct 06;18(10):e268 [FREE Full text] [doi: [10.2196/jmir.5916](https://doi.org/10.2196/jmir.5916)] [Medline: [27713114](https://pubmed.ncbi.nlm.nih.gov/27713114/)]
170. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. *Scand J Med Sci Sports* 2018 Jul;28(7):1818-1827. [doi: [10.1111/sms.13074](https://doi.org/10.1111/sms.13074)] [Medline: [29460319](https://pubmed.ncbi.nlm.nih.gov/29460319/)]
171. De Ridder R, De Blaiser C. Activity trackers are not valid for step count registration when walking with crutches. *Gait Posture* 2019 May;70:30-32. [doi: [10.1016/j.gaitpost.2019.02.009](https://doi.org/10.1016/j.gaitpost.2019.02.009)] [Medline: [30798092](https://pubmed.ncbi.nlm.nih.gov/30798092/)]
172. Boeselt T, Spielmanns M, Nell C, Storre JH, Windisch W, Magerhans L, et al. Validity and Usability of Physical Activity Monitoring in Patients with Chronic Obstructive Pulmonary Disease (COPD). *PLoS One* 2016;11(6):e0157229 [FREE Full text] [doi: [10.1371/journal.pone.0157229](https://doi.org/10.1371/journal.pone.0157229)] [Medline: [27305105](https://pubmed.ncbi.nlm.nih.gov/27305105/)]
173. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *J Sports Sci* 2015 Dec 17;33(10):991-1000. [doi: [10.1080/02640414.2014.981848](https://doi.org/10.1080/02640414.2014.981848)] [Medline: [25517396](https://pubmed.ncbi.nlm.nih.gov/25517396/)]
174. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. *Ann Transl Med* 2016 Aug;4(15):278-278 [FREE Full text] [doi: [10.21037/atm.2016.07.16](https://doi.org/10.21037/atm.2016.07.16)] [Medline: [27570772](https://pubmed.ncbi.nlm.nih.gov/27570772/)]
175. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. *Int J Med Inform* 2017 Jun;102:87-92. [doi: [10.1016/j.ijmedinf.2017.03.008](https://doi.org/10.1016/j.ijmedinf.2017.03.008)] [Medline: [28495352](https://pubmed.ncbi.nlm.nih.gov/28495352/)]
176. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of Wrist-Worn Heart Rate Monitors. *JAMA Cardiol* 2017 Jan 01;2(1):104-106. [doi: [10.1001/jamacardio.2016.3340](https://doi.org/10.1001/jamacardio.2016.3340)] [Medline: [27732703](https://pubmed.ncbi.nlm.nih.gov/27732703/)]
177. Roos L, Taube W, Beeler N, Wyss T. Validity of sports watches when estimating energy expenditure during running. *BMC Sports Sci Med Rehabil* 2017 Dec 20;9(1):22 [FREE Full text] [doi: [10.1186/s13102-017-0089-6](https://doi.org/10.1186/s13102-017-0089-6)] [Medline: [29296281](https://pubmed.ncbi.nlm.nih.gov/29296281/)]
178. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015 Dec 18;12:159 [FREE Full text] [doi: [10.1186/s12966-015-0314-1](https://doi.org/10.1186/s12966-015-0314-1)] [Medline: [26684758](https://pubmed.ncbi.nlm.nih.gov/26684758/)]
179. Gerrior S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. *Prev Chronic Dis* 2006 Oct;3(4):A129 [FREE Full text] [Medline: [16978504](https://pubmed.ncbi.nlm.nih.gov/16978504/)]
180. Open mHealth. URL: <https://www.openmhealth.org/> [accessed 2020-08-17]
181. Ralls J. Fitbit to Be Acquired by Google. Fitbit. 2019. URL: <https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-to-Be-Acquired-by-Google/default.aspx> [accessed 2019-11-01]

182. Osterloh R. Helping more people with wearables: Google to acquire Fitbit Internet. Google Blog. URL: <https://blog.google/products/hardware/agreement-with-fitbit? ga=2.109995341.918473813.1572613323-1996097189.1566566630> [accessed 2020-08-17]
183. Morrison A, Polisen J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012 Apr;28(2):138-144. [doi: [10.1017/S0266462312000086](https://doi.org/10.1017/S0266462312000086)] [Medline: [22559755](https://pubmed.ncbi.nlm.nih.gov/22559755/)]
184. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002 Feb;31(1):115-123. [doi: [10.1093/ije/31.1.115](https://doi.org/10.1093/ije/31.1.115)] [Medline: [11914306](https://pubmed.ncbi.nlm.nih.gov/11914306/)]

## Abbreviations

**API:** application program interface

**COSMIN:** Consensus-Based Standards for the Selection of Health Status Measurement Instruments

**MAPE:** mean absolute percentage error

**MPE:** mean percentage error

*Edited by G Eysenbach; submitted 30.03.20; peer-reviewed by D Lott, L Becker; comments to author 12.06.20; revised version received 22.06.20; accepted 25.06.20; published 08.09.20*

*Please cite as:*

*Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review*

*JMIR Mhealth Uhealth* 2020;8(9):e18694

URL: <http://mhealth.jmir.org/2020/9/e18694/>

doi: [10.2196/18694](https://doi.org/10.2196/18694)

PMID: [32897239](https://pubmed.ncbi.nlm.nih.gov/32897239/)

©Daniel Fuller, Emily Colwell, Jonathan Low, Kassia Orychock, Melissa Ann Tobin, Bo Simango, Richard Buote, Desiree Van Heerden, Hui Luan, Kimberley Cullen, Logan Slade, Nathan G A Taylor. Originally published in JMIR mHealth and uHealth (<http://mhealth.jmir.org>), 08.09.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.