

Review

Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature

Federico Germini^{1,2}, MSc, MD; Noella Noronha^{1,3}, MSc; Victoria Borg Debono¹, MSc, PhD; Binu Abraham Philip¹, MSc; Drashti Pete¹, MPH; Tamara Navarro¹, MLIS, MEd; Arun Keepanasseril^{1,2}, MSc, MDS; Sameer Parpia^{1,4}, PhD; Kerstin de Wit^{1,5}, MBChB, MSc, MD; Alfonso Iorio^{1,2}, MD, PhD

¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

²Department of Medicine, McMaster University, Hamilton, ON, Canada

³School of Interdisciplinary Sciences, McMaster University, Hamilton, ON, Canada

⁴Department of Oncology, McMaster University, Hamilton, ON, Canada

⁵Department of Emergency Medicine, Queen's University, Kingston, ON, Canada

Corresponding Author:

Noella Noronha, MSc

Department of Health Research Methods, Evidence, and Impact

McMaster University

1280 Main Street West

Hamilton, ON, L8S 4L8

Canada

Phone: 1 9055259140 ext 22069

Email: noronn@mcmaster.ca

Abstract

Background: Numerous wrist-wearable devices to measure physical activity are currently available, but there is a need to unify the evidence on how they compare in terms of acceptability and accuracy.

Objective: The aim of this study is to perform a systematic review of the literature to assess the accuracy and acceptability (willingness to use the device for the task it is designed to support) of wrist-wearable activity trackers.

Methods: We searched MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, and SPORTDiscus for studies measuring physical activity in the general population using wrist-wearable activity trackers. We screened articles for inclusion and, for the included studies, reported data on the studies' setting and population, outcome measured, and risk of bias.

Results: A total of 65 articles were included in our review. Accuracy was assessed for 14 different outcomes, which can be classified in the following categories: count of specific activities (including step counts), time spent being active, intensity of physical activity (including energy expenditure), heart rate, distance, and speed. Substantial clinical heterogeneity did not allow us to perform a meta-analysis of the results. The outcomes assessed most frequently were step counts, heart rate, and energy expenditure. For step counts, the Fitbit Charge (or the Fitbit Charge HR) had a mean absolute percentage error (MAPE) <25% across 20 studies. For heart rate, the Apple Watch had a MAPE <10% in 2 studies. For energy expenditure, the MAPE was >30% for all the brands, showing poor accuracy across devices. Acceptability was most frequently measured through data availability and wearing time. Data availability was ≥75% for the Fitbit Charge HR, Fitbit Flex 2, and Garmin Vivofit. The wearing time was 89% for both the GENEActiv and Nike FuelBand.

Conclusions: The Fitbit Charge and Fitbit Charge HR were consistently shown to have a good accuracy for step counts and the Apple Watch for measuring heart rate. None of the tested devices proved to be accurate in measuring energy expenditure. Efforts should be made to reduce the heterogeneity among studies.

(*J Med Internet Res* 2022;24(1):e30791) doi: [10.2196/30791](https://doi.org/10.2196/30791)

KEYWORDS

diagnosis; measurement; wrist-wearable devices; mobile phone

Introduction

Background

Tracking, measuring, and documenting one's physical activity can be a way of monitoring and encouraging oneself to participate in daily physical activity; increased activity is thought to translate into important positive health outcomes, both physical and mental [1]. In the past, most physical activity tracking was done manually by oneself or an external assessor, through records, logbooks, or using questionnaires. These are indirect methods to quantify physical activity, meaning that they do not measure movement directly as it occurs [2]. The main disadvantages of such methods are the administrative burden on either the self-assessor or the external assessor and the potential imprecision because of recall bias [2,3]. Direct methods to assess physical activity [2], such as accelerometers or pedometers that digitally record movement, are preferred because they eliminate recall bias and are convenient. This process of activity tracking has become automated, accessible, and digitized with wearable tracking technology such as wristband sensors and smartwatches that can be linked to computer apps on other devices such as smartphones, tablets, and PCs. When data are uploaded to these devices, users can review their physical activity log and potentially use this feedback to make behavior changes with regard to physical activity.

The ideal device should be acceptable to the end user, affordable, easy to use, and accurate in measuring physical activity. Accuracy can be defined as the closeness of the measured value to the actual value. Accuracy can be calculated using measures of agreement, sensitivity and specificity, receiver operating characteristic curves, or absolute and percentage differences [4]. Agreement can be defined as "the degree of concordance between two or more sets of measurements" [5]. It can be measured as percentage agreement, that is, the percentage of cases in which 2 methods of measurements of the same variable lead to classification in the same category. Another example of methods of calculating agreement is the κ statistic, which measures agreement beyond chance [6]. Sensitivity and specificity are the true positive and true negative proportions, respectively. These proportions are calculated using the measurement method that we are evaluating as the index test and another method, known to be accurate, as the reference standard [4]. Receiver operating characteristic curves are obtained plotting the sensitivity versus the complement of specificity and can be used to find optimal cutoff points for the index test. Absolute and percentage differences are used to determine how far the index test measurement is from the reference standard or their average [4]. Acceptability can be widely defined as "the demonstrable willingness within a user group to employ information technology for the task it is designed to support" [7]. It can be assessed qualitatively (eg, through questionnaires or interviews) or quantitatively (eg, percentage of the time during which the device is worn or the data are available or measured using ad hoc scales). On the basis of a 2019 review, acceptability or acceptance of wrist-wearable activity-tracking devices is dependent on the type of user and context of use [8]. This same review indicates that research on

accuracy has not kept up with the plethora of wearable physical activity-tracking devices in the market [8]. This may be because of the rapidly changing landscape as companies continue to upgrade models with different technical specifications and features. The purpose of this systematic review is to assess the acceptability and accuracy of these wrist-wearable activity-tracking devices through a focused in-depth review of primary studies assessing these 2 characteristics.

Objectives

The first objective of this systematic review is to assess the accuracy of wrist-wearable activity-tracking devices for measuring physical activity.

The second objective is to assess the acceptability of wrist-wearable activity-tracking devices for measuring physical activity.

Methods

The methods used for this systematic review have been registered in the PROSPERO database (CRD42019137420).

Search Strategy

The databases searched were MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, and SPORTDiscus from inception to May 28, 2019. Search strategies were developed to retrieve content on wearable activity trackers and on their accuracy and reproducibility of results. We used search terms, including *Wearable device* and *Fitness tracker*, to identify studies on the use of a consumer-based wearable activity tracker, whereas terms such as *data accuracy* and *reproducibility of results* were included to bring in content focused on activity tracker validation. The search strategy is available on the web in the PROSPERO record. A snowball search was conducted by checking the references of relevant studies and systematic reviews on this topic that were identified in our original search.

Selection of Studies

For the acceptability objective, the population was the general population, without sex or age restrictions. The intervention was the use of a wrist-wearable activity tracker. The outcome was any quantitative measure of acceptability, including wearing time, data availability, and questionnaires to assess acceptability.

For the accuracy objective, the population was again the general population, the index test had to be a wrist-wearable activity tracker, and the reference standard could be another device or any method used to measure physical activity, including questionnaires and direct observation. The outcome could be any measure of physical activity, including but not limited to step count, heart rate, distance, speed, activity count, activity time, and intensity of physical activity.

For both objectives, this review examined both research-grade devices (activity trackers available only for research purposes) and commercial devices (those available to the general public). The included studies were limited to the community-based everyday-life setting. Laboratory tests such as research studies were included as long as everyday settings were reproduced, thereby excluding patients who were institutionalized and those

who were hospitalized. We set no restrictions on the length of observation for the original studies.

The exclusion criteria included the following: device not worn on the wrist, studies measuring sleep, and studies on patients who were institutionalized or hospitalized.

All studies reporting primary data were considered for inclusion, with the exception of case reports and case series.

Using the aforementioned inclusion and exclusion criteria and a piloted form, we initially screened for inclusion from the titles and abstracts of the retrieved articles, using the web-based software Rayyan (Rayyan Systems Inc) [9]. Subsequently, we screened the full texts of the studies identified as potentially eligible from the title and abstract screening for selection.

Data Extraction and Risk-of-Bias Assessment

Data were extracted to an Excel (Microsoft Corp) file. The data extraction form was based on a previous publication on the same topic [8] and adapted to the needs of this review. The following data were extracted: general study information: first author's name, publication year, type of study (prospective vs retrospective and observational vs interventional), duration of follow-up (in days), and setting (laboratory vs field); characteristics of the population: number of participants, underlying health condition (eg, healthy participants, people with severe obesity, and chronic joint pain), gender, and age distribution (mean and SD or median and minimum–maximum or first and third quartiles); measures of accuracy: step count, distance, speed, heart rate, activity count, time spent being active, and intensity of physical activity; and acceptability of the device, including but not limited to data availability, wearing time, ease of use. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies, version 2, tool [10]. This tool guides the assessment of the risk of bias in diagnostic accuracy studies in 4 domains: patient selection, index test, reference standard, and flow and timing. We rated the risk of bias in each domain as *High*, *Probably high*, *Probably low*, and *Low*. When necessary, the study authors were contacted for additional information.

Throughout title and abstract and full-text screening and the data extraction, each step was performed in duplicate with 2 reviewers (NN and BAP) deciding independently on inclusion or exclusion and, if needed, later having a discussion with another author to make a final decision. Disagreements were solved through discussion and, when needed, with the intervention of a third reviewer (FG, VBD, or DP). The reviewers were trained with calibration exercises until an adequate performance was achieved for each of these steps.

Diagnostic Accuracy Measures

When available, we extracted the mean absolute percentage error (MAPE) or the mean percentage error. When these were

not available, we extracted other measures in the following order of priority: mean difference, mean bias (Bland–Altman), accuracy determined through intraclass correlation coefficient, and correlation coefficient (Pearson or Spearman). When the outcome was dichotomized and sensitivity and specificity were calculated, we reported on these values. When available, we reported measures of variability or 95% CIs for all the aforementioned measures. The formulas used for calculating the MAPE, mean percentage error, mean difference, and mean bias are reported in [Multimedia Appendix 1](#) [11-75].

Synthesis of Results

Because of the significant heterogeneity observed in the studies' populations, settings, devices assessed, reference standards, outcomes assessed, and the outcome measures reported, we decided not to perform a quantitative synthesis and have provided a narrative synthesis of the results for both the objectives. For the accuracy objective, given the high number of studies retrieved, we summarized results only for devices that were included in at least two studies reporting the same outcome. All the remaining results are reported in [Multimedia Appendix 1](#).

Ethics Approval

This systematic review was based on published data and therefore did not require a submission to a research ethics board.

Availability of Data and Materials

Most of the data that support the findings of this study are available in [Multimedia Appendix 1](#). A guide on how to use the database provided in [Multimedia Appendix 1](#) can be found in [Multimedia Appendix 2](#). The full data set can be made available upon reasonable request.

Code Availability

This is not applicable to this systematic review because no quantitative data synthesis was performed.

Results

Overview

The search identified 1633 records (1614, 98.84%, after the removal of duplicates). The study flow diagram is presented in [Figure 1](#). After screening the full texts of 398 articles, 65 (16.3%) were included in the systematic review. The characteristics of the included studies are summarized in [Table 1](#) and [Multimedia Appendix 3](#) [11-67] for the accuracy objective and [Table 2](#) for the acceptability objective. All the included studies were single-center studies, with a prospective, observational design. The complete results for accuracy and acceptability have been reported in [Multimedia Appendix 1](#).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 study flow diagram.

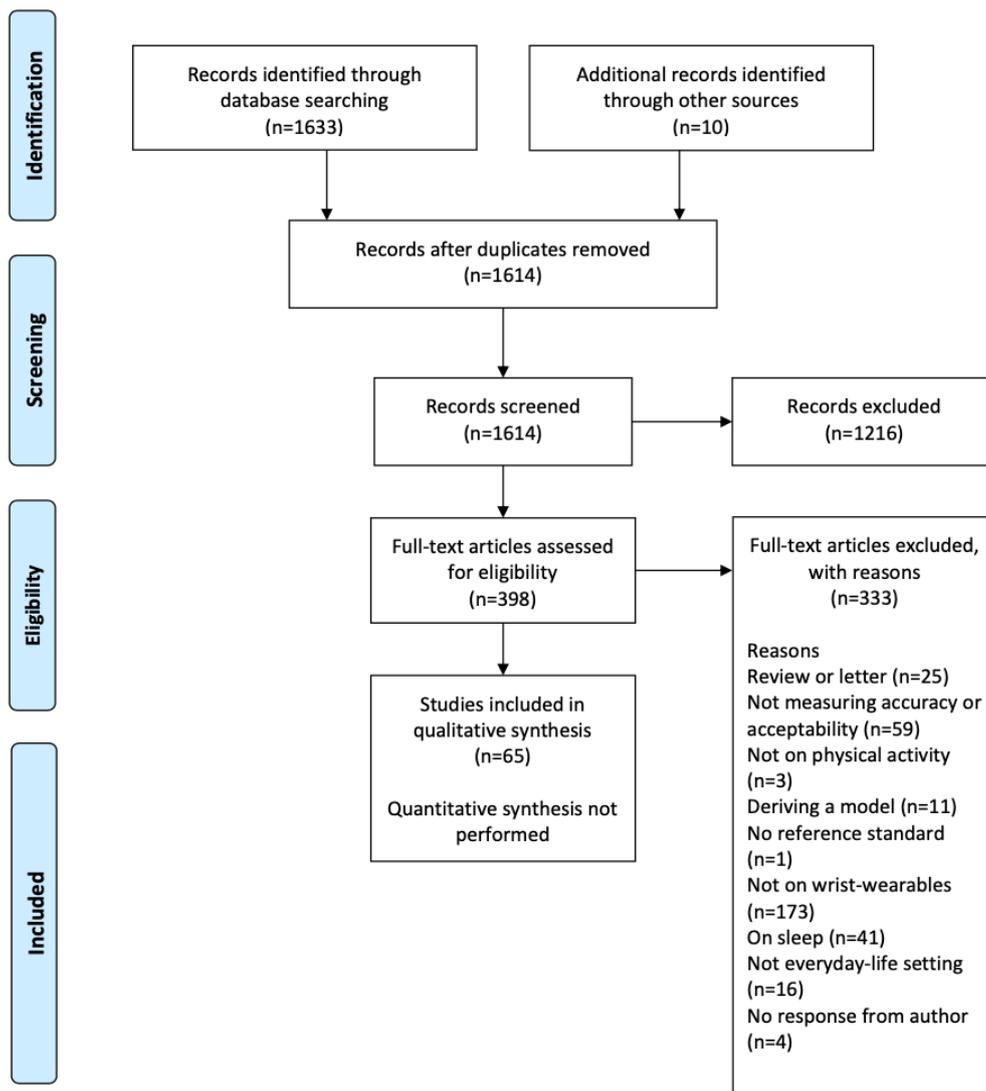


Table 1. Characteristics of the studies reporting on accuracy (N=57).

First author, year	Setting	FUP ^a time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Alharbi [11], 2016	Laboratory	<1	48	66 (7)	48	Patients undergoing cardiac rehabilitation	Step count and MVPA ^b	Fitbit Flex
Alsubheen [12], 2016	Field	5	13	40 (12)	38	Healthy	Step count and energy expenditure	Garmin Vivofit
An [13], 2017	Laboratory and field	1	35	31 (12)	51	Healthy	Step count	Fitbit Flex, Garmin Vivofit, Polar Loop, Basis B1 Band, Misfit Shine, Jawbone UP24, and Nike FuelBand SE
An [14], 2017	Field	<1	62	24 (5)	40	Healthy	Active time	ActiGraph GT3X
Blondeel [15], 2018	Field	14	8	65 (8)	25	COPD ^c	Step count	Fitbit Alta
Boeselt [16], 2016	Field	3	20	66 (7)	15	COPD	Step count, energy expenditure, and MVPA	Polar A300
Bruder [17], 2018	Laboratory and field	395	32	— ^d	—	Rehabilitation after radial fracture	Activity count	ActivPAL
Bulathsinghala [18], 2014	Laboratory	1	20	70 (10)	—	COPD	Physical activity intensity	ActiGraph GT3X+
Burton [19], 2018	Laboratory	<1	31	74 (6)	65	Healthy older adults	Step count	Fitbit Flex and Fitbit Charge HR
Choi [20], 2010	Laboratory	<1	76	13 (2)	62	Healthy	Energy expenditure	ActiGraph GT1M
Chow [21], 2017	Laboratory	1.5	31	24 (5)	39	Healthy	Step count	ActiGraph wGT3xBT-BT, Fitbit Flex, Fitbit Charge HR, and Jawbone UP24
Chowdhury [22], 2017	Laboratory and field	2	30	27 (6)	50	Healthy	Energy expenditure	Microsoft Band; Apple Watch, series not specified; Jawbone Up24; and Fitbit Charge
Cohen [23], 2010	Laboratory and field	3	57	70 (10)	—	COPD	Speed	ActiGraph Mini Motionlogger
Compagnat [24], 2018	Laboratory	<1	46	65 (13)	—	Stroke	Energy expenditure	ActiGraph GT3X+
Dondzila [25], 2018	Laboratory and field	3-62	40	22 (2)	58	Healthy	Step count, energy expenditure, and heart rate	Fitbit Charge HR and Mio Fuse
Dooley [26], 2017	Laboratory	1	62	23 (4)	58	Healthy	Heart rate and energy expenditure	Apple Watch, series not specified; Fitbit Charge HR; and Garmin Forerunner 225
Durkalec-Michalski [27], 2013	Laboratory	2	20	26 (5)	55	Healthy	Energy expenditure	ActiGraph GT1M
Falgoust [28], 2018	Laboratory	1	30	—	—	Healthy	Step count	Fitbit Charge HR, Fitbit Surge, and Garmin Vivoactive HR

First author, year	Setting	FUP ^a time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Ferguson [29], 2015	Field	2	21	33 (10)	52	Healthy	Step count, MV-PA, and energy expenditure	Nike FuelBand, Misfit Shine, and Jawbone UP
Gaz [30], 2018	Laboratory and field	<1	32	36 (8)	69	Healthy	Step count and distance	Fitbit Charge HR; Apple Watch, series not specified; Garmin Vivofit 2; and Jawbone UP2
Gillinov [31], 2017	Laboratory	<1	50	38 (12)	54	Healthy	Heart rate	Garmin Forerunner 235; TomTom Spark; Apple Watch, series not specified; and Fitbit Blaze
Gironda [32], 2007	Laboratory	<1	3	43 ^e	31	Pain syndromes	Activity count	Actiwatch Score
Hargens [33], 2017	Laboratory and field	7	21	31 ^e	68	Healthy	MVPA, energy expenditure, and step count	Fitbit Charge
Hernandez-Vicente [34], 2016	Field	7	18	21 (1)	50	Healthy	Energy expenditure, vigorous active time, active time, and step count	Polar V800
Huang [35], 2016	Laboratory	1	40	24 (3)	25	Healthy	Step count and distance	Jawbone UP24, Garmin Vivofit, Fitbit Flex, and Nike FuelBand
Imboden [36], 2018	Laboratory	<1	30	49 (19)	50	Healthy	Energy expenditure, MVPA, and step count	Fitbit Flex, Jawbone Up24, and Fitbit Flex
Jo [37], 2016	Laboratory	<1	24	25 ^e	50	Healthy	Heart rate	Basis Peak K and Fitbit Charge
Jones [38], 2018	Laboratory	118	30	33 ^e	—	Healthy	Step count	Fitbit Flex
Kaewkannate [39], 2016	Laboratory	<1	7	31 (0)	14	Healthy	Step count	Fitbit Flex, Jawbone UP24, Withings Pulse, and Misfit Shine
Lamont [40], 2018	Laboratory	<1	33	67 (8)	64	Parkinson disease	Step count	Garmin Vivosmart HR and Fitbit Charge HR
Lauritzen [41], 2013	Laboratory and field	<1	18	—	56	Older adults	Step count	Fitbit Ultra
Lawinger [42], 2015	Laboratory	<1	30	26 (6)	70	Healthy	Activity count	ActiGraph GT3X+
Lemmens [43], 2018	Laboratory	<1	40	31 (5)	100	Parkinson disease	Energy expenditure	Philips optical heart rate monitor
Magistro [44], 2018	Laboratory	<1	40	74 (7)	60	Healthy	Step count	ADAMO Care Watch
Mandigout [45], 2017	Laboratory	<1	24	68 (14)	60	Stroke	Energy expenditure	Actical and ActiGraph GTX
Manning [46], 2016	Laboratory	<1	9	15 (1)	—	Severe obesity	Step count	Fitbit One, Fitbit Flex, Fitbit Zip, ActiGraph GT3x+, and Jawbone UP

First author, year	Setting	FUP ^a time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Montoye [47], 2017	Laboratory	<1	30	24 (1)	47	Healthy	Step count, energy expenditure, and heart rate	Fitbit Charge HR
Powierza [48], 2017	Field	<1	22	22 (2)	55	Healthy	Heart rate	Fitbit Charge
Price [49], 2017	Laboratory	<1	14	23 ^e	21	Healthy	Energy expenditure	Fitbit One, Garmin Vivofit, and Jawbone UP
Redenius [50], 2019	Laboratory	4	65	42 (12)	72	Healthy	MVPA	Fitbit Flex
Reid [51], 2017	Field	4	22	21 (2)	100	Healthy	MVPA and step count	Fitbit Flex
Roos [52], 2017	Laboratory	2	20	24 (2)	40	Runners	Energy expenditure	Suunto Ambit, Garmin Forerunner 920XT, and Polar V800
Schaffer [53], 2017	Laboratory	<1	24	54 (13)	42	Stroke	Step count	Garmin Vivofit
Scott [54], 2017	Field	7	89	—	54	Healthy	Daily mean activity and MVPA	GENEActiv
Semanik [55], 2020	Laboratory	7	35	52 ^e	69	Chronic joint pain	MVPA	Fitbit Flex
Sirard [56], 2017	Laboratory and field	7	14	9 (2)	50	Healthy	Energy expenditure, MVPA, and step count	Movband and Sqord
St-Laurent [57], 2018	Laboratory	7	16	33 (4)	100	Pregnant	Step count and MVPA	Fitbit Flex
Stackpool [58], 2013	Laboratory	<1	20	22 (1)	50	Healthy	Step count and energy expenditure	Jawbone UP, Nike FuelBand, Fitbit Ultra, and Adidas miCoach
Stiles [59], 2013	Laboratory	1	10 ^e	39 (6)	100	Healthy premenopausal women	Loading rate (BW ^f /s)	GENEActiv and ActiGraph GT3X+
Støve [60], 2019	Laboratory	<1	29	29 (9)	41	Healthy	Heart rate	Garmin Forerunner
Tam [61], 2018	Laboratory	<1	30	32 (9)	50	Healthy	Step count	Fitbit Charge HR and Xiaomi Mi Band 2
Thomson [62], 2019	Laboratory	<1	30	24 (3)	50	Healthy	Heart rate	Apple Watch, series not specified; and Fitbit Charge HR2
Wahl [63], 2017	Laboratory	<1	20	25 (3)	50	Healthy	Step count, energy expenditure, and distance	Polar Loop, Beurer AS80, Fitbit Charge HR, Fitbit Charge, Bodymedia Sensewear, Garmin Vivofit, Garmin Vivosmart, Garmin Vivoactive, Garmin Forerunner 920XT, Xiaomi Mi Band, and Withings Pulse
Wallen [64], 2016	Laboratory	<1	22	24 (6)	50	Healthy	Heart rate, energy expenditure, and step count	Apple Watch, series not specified; Samsung Gear S; Mio Alpha; and Fitbit Charge

First author, year	Setting	FUP ^a time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Wang [65], 2017	Laboratory	<1	9	22 (1)	44	Healthy	Step count	Huawei B1, Xiaomi Mi Band, Fitbit Charge, Polar Loop, Garmin Vivofit 2, Misfit Shine, and Jawbone UP
Woodman, 2017 [66]	Laboratory	<1	28	25 (4)	29	Healthy	Energy expenditure	Garmin Vivofit, Withings Pulse, and Basis Peak
Zhang [67], 2012	Laboratory	1	60	49 (7)	62	Healthy	Activity classification (sedentary, household, walking, and running)	GENEActiv

^aFUP: follow-up.

^bMVPA: moderate- to vigorous-intensity physical activity.

^cCOPD: chronic obstructive pulmonary disease.

^dNot available.

^eSD not reported.

^fBW: body weight.

Table 2. Characteristics of the studies reporting on acceptability (N=11).

First author, year	Setting	FUP ^a time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome assessed	Device brand and model
Boeselt [16], 2016	Laboratory and field	7	20	66 (7)	15	COPD ^b	Ease of use and other characteristics	Polar A300
Deka [68], 2018	Field	5	46	65 (12)	67	CHF ^c	Data availability	Fitbit Charge HR
Farina [69], 2019	Field	2	26; 26	80 (6); 76 (6)	39; 73	Dementia; caregivers of patients with dementia	Wearing time	GENEActiv
Fisher [70], 2016	Field	7	34	69 ^d	— ^e	Parkinson disease	Ease of use and other characteristics	AX3 data logger
Kaewkannate [39], 2016	Field	<1	7	31 (0)	14	Healthy	Ease of use and other characteristics	Fitbit Flex, Jawbone UP24, Withings Pulse, and Misfit Shine
Lahti [71], 2017	Laboratory	120	40	—	—	Schizophrenia	Data availability	Garmin Vivofit
Marcoux [72], 2019	Field	46	20	73 (7)	20	Idiopathic pulmonary fibrosis	Data availability	Fitbit Flex 2
Naslund [73], 2015	Field	80-133	5	48 (9)	90	Serious mental illness	Wearing time	Nike Fuel-Band
Speier [74], 2018	Laboratory	90	186	—	—	Coronary artery disease	Wearing time	Fitbit Charge HR2
St-Laurent [57], 2018	Laboratory	1	16	33 (4)	100	Pregnant	Ease of use and other characteristics	Fitbit Flex
Rowlands [75], 2018	Field	425	1724	13 (1)	100	Healthy	Data availability	GENEActiv

^aFUP: follow-up.

^bCOPD: chronic obstructive pulmonary disease.

^cCHF: congestive heart failure.

^dSD not reported.

^eNot available.

Accuracy

The accuracy of wrist-wearable activity trackers was assessed in 57 studies on 72 devices from 29 brands. Step count, heart rate, and energy expenditure (EE) were the most commonly assessed outcomes in the appraised literature. The results of

these outcomes are summarized in [Figure 2](#) (icons by Nikhil Bapna, Yoyon Pujiyono, Chintuza, Gregor Cresnar, Andrejs Kirma, and Yigit Pinarbasi from the Noun Project [76]), in which we have highlighted the standout device for the most frequently reported outcomes.

Figure 2. Summary of the results for the main accuracy outcomes. MAPE: mean absolute percentage error.

Outcome	Number of studies	Number of brands	Number of devices
 Step counts	31	29	72
	Fitbit Charge (±HR): MAPE<25% in 20 (65%) studies		
 Heart rate	9	7	15
	Apple Watch: MAPE<10% in 2 (22%) studies		
 Energy expenditure	22	22	36
	MAPE>30% for all the brands		

Step Counts

A total of 31 studies on 72 devices from 29 brands reported data on step counts. The reference standards used were manual count (directly observed or on video, usually with the help of a tally counter) or automated count through video analysis, an activity tracker (8 different devices), or a photoelectric cell.

The *ActiGraph wGT3xBT-BT*, tested against manual count, showed a mean percentage error of -41.7% (SD 13.5%) [21]. The *ActiGraph GT3x+* showed no statistically significant correlation with the same reference standard [46].

The *Apple Watch* (series not specified) was evaluated in 6% (2/31) of studies using manual count as the reference standard [30,64]. The mean difference between the device and the manual count varied from -47 (SD 470) steps to 39.44 (SD 151.81) steps in different walking conditions.

For the *Fitbit Alta*, the mean step count was 773 (SD 829) higher ($P=.009$) than the one obtained with the reference standard, an accelerometer [15]. For the *Fitbit Charge*, the mean difference was -59 (SD 704) steps compared with direct observation [64]. The MAPE for the same device ranged from -4.4% to 20.7%, using different automated step count methods as the reference standard [33,63,65]. The *Fitbit Charge HR* was assessed in 29% (9/31) of studies, using direct observation [19,21,28,30,61] or an automated method of step count as the reference standard

[25,40,47,63]. The MAPE ranged from -12.7% to 24.1%. The accuracy of the *Fitbit Flex* in measuring steps was assessed in 35% (11/31) of studies, using manual count [13,19,21,35,36,38,39,46] or an ActiGraph device [11,51,57] as the reference standard. The mean percentage error ranged from -23% to 13%. For the *Fitbit One* and *Fitbit Zip*, no statistically significant correlation was found in step counting using direct observation as the reference standard [46]. The correlation coefficient was not reported. For the *Fitbit Surge*, the mean difference compared with direct observation was -86.0 steps ($P=.004$) [28]. For the *Fitbit Ultra*, the MAPE was 99.6% (SD 0.8%) [41] and the Pearson correlation coefficient against manual count ranged from 0.44 to 0.99 in different exercise conditions [58].

The accuracy of the *Garmin Vivofit* was assessed in 16% (5/31) of studies [12,13,35,53,63], with a MAPE ranging from -41% to 18% [13,53,63]. For the *Vivofit 2*, a study reported a MAPE of 4% [65] and another study reported a mean difference ranging from 5.09 (SD 8.38) steps to 98.06 (SD 137.49) steps in different walking conditions (over a maximum distance of 1.6 km) [30].

In a study by Wahl et al [63], the MAPE against automated step counting using a photoelectric cell as the reference standard, in different exercise types and conditions, ranged from -2.7% to 1.5% for the *Garmin Forerunner 920XT*, from -1.5% to 0.6% for the *Garmin Vivoactiv*, and from -1.1% to -0.3% for the *Garmin Vivosmart* [63]. For the *Garmin Vivoactive HR*, the

mean difference against manual step count was -19.7 steps ($P=.03$) [28]. For the *Garmin Vivosmart HR*, the mean difference ranged from -39.7 (SD 54.9) steps to 5.4 (SD 5.8) steps for different walking speeds and locations (outdoor vs indoor) over a total of 111-686 steps [40].

For the *Jawbone UP*, the MAPE was -6.73% in a study [65] and the mean absolute difference 806 over an average of 9959 steps in another study [29]. For the *Jawbone UP2*, the mean difference ranged from 16.19 (SD 29.14) steps to 64 (SD 66.32) steps for different walking conditions over a maximum distance of 1.6 km [30]. For the *Jawbone UP24*, the mean percentage error ranged from -28% to -0.8% [21,35,36].

For the *Misfit Shine*, the MAPE ranged from -13% to 23% [13,65].

For the *Mio Fuse*, the MAPE ranged from -5% to -16% at different treadmill speeds [25], whereas in another study, the mean percentage error was $<5\%$ for the *Xiaomi Mi Band 2* [61].

For the *Nike FuelBand*, the mean percentage error ranged from -34.3% (SD 26.8%) to -16.7% (SD 16.5%) [35], whereas for the *FuelBand SE*, the MAPE ranged from 10.2% to 45.0% [13].

The MAPE for the *Polar Loop* ranged from -13% to 27% in 3 studies [13,63,65]. Regarding 2 other devices from *Polar*, for the *A300*, a Pearson correlation coefficient of 0.96 ($P<.01$) [16] was reported, whereas for the *V800*, the Bland-Altman bias was equal to 2487 (SD 2293) steps per day over a mean 10,832 (SD 4578) steps per day measured with the reference standard [34].

For the *Withings Pulse*, the MAPE for step count ranged from -16.0% to -0.4% [63] and the accuracy from 97.2% to 99.9% [39]. All the remaining devices were only used in 1 study each, and the results are reported in [Multimedia Appendix 1](#).

Heart Rate

A total of 9 studies on 15 devices from 7 brands evaluated the accuracy of activity-tracking devices to measure the participants' heart rates. The reference standards used were electrocardiography, pulse oximetry, or another activity tracker (4 different devices).

For the *Apple Watch*, the MAPE for measuring heart rate ranged from 1% (SD $\sim 1\%$) to 7% (SD $\sim 11\%$) [26,31].

In the *Fitbit* family of devices, for the *Fitbit Charge*, the mean bias estimated with the Bland-Altman method ranged from -6 (SD 10) bpm to -9 (SD 8) bpm [37,48,64]. For the *Fitbit Charge HR* or *Fitbit ChargeHR2*, the MAPE for measuring heart rate ranged from 2.4% (SD $\sim 1.5\%$) to 17% (SD $\sim 20\%$) [26,47,62]. For the *Fitbit Blaze*, the MAPE ranged from 6% (SD 6%) to 16% (SD 18%) for different activities [31].

Active Time: Time Spent in Moderate- to Vigorous-Intensity Physical Activity and Other Outcomes

A total of 13 studies on 11 devices from 8 brands reported on the time spent being active, most frequently defined as the time spent in moderate- to vigorous-intensity physical activity (MVPA; 11 studies), expressed in minutes per day. The reference standard for MVPA was another activity tracker (3

different devices). Other outcomes were time spent being active (standing+walking+running), time spent running, or time spent on different types of physical activity, with each of these outcomes being reported in only 1 study.

For the *Fitbit Flex*, the MAPE for measuring the time spent in MVPA varied from 7% (SD 6%) to 74% (SD 13%) [50] and the mean percentage error ranged from -65% to 10% [11,36]. All the other devices were only used in 1 study each, and the results are reported in [Multimedia Appendix 1](#).

Intensity of Activity: EE and Other Outcomes

A total of 24 studies on 42 devices from 23 brands focused on measuring the intensity of physical activity. The most frequent measure of intensity was EE, expressed as kcal, evaluated in 92% (22/24) of studies. The less frequent measures of intensity included loading rate and the classification of physical activity (sedentary, household, walking, and running). For EE, the reference standard used most commonly was indirect calorimetry (6 different instruments). Less common reference standards included EE estimated with other wearable activity trackers (5 different devices), estimated based on the treadmill settings, or direct room calorimetry.

Among the *ActiGraph* family, the mean percentage difference in the EE compared with the reference standard in people with previous stroke was 3% for walking participants and 47% for participants with wheelchair using the *ActiGraph GT3X+* [24]. The Spearman correlation coefficient was 0.08 ($P=.71$) if worn on the plegic side and 0.20 ($P=.34$) if worn on the nonplegic side with the *ActiGraph GTX* [45]. Using the *ActiGraph GTIM*, the mean percentage difference was 0.5% (SD 8.0%) in a study [20], whereas another study found that the device overestimated EE at moderate intensity by 60% and underestimated EE by 40% at vigorous intensity while being 86% accurate in measuring EE at light intensity [27].

For the *Apple Watch*, the MAPE for EE ranged from 15% (SD 10%) to 211% (SD $\sim 96\%$) [22,26].

In the *Fitbit* family, the MAPE from the *Charge* model ranged from -4.5% to 75.0% in different studies [22,33,63] and from -12% to 89% for the *Charge HR* [25,26,47,63]. For the *Fitbit Flex*, a mean percentage bias of -13% was reported [36]. For the *Fitbit One*, a study reported a mean bias of 2.91 (SD 4.35) kcal per minute [49], whereas for the *Fitbit Ultra*, the Pearson correlation coefficient ranged from 0.24 to 0.67 for different physical activities [58].

Among the devices from *Garmin*, the MAPE for EE ranged from -21% to 45% for the *Vivofit* [63,66], from -2% to -36% for the *Vivosmart* [63], and from 5% to 37% for the *Vivoactive* [63].

For the *Garmin Forerunner*, the MAPE ranged from -27% to 49% for the model 920XT [52,63] and from 31% (SD $\sim 26\%$) to 155% (SD $\sim 164\%$) for the model 225 [26].

In the *Polar* family, the MAPE for EE ranged from 10% to 40% for the *V800* model [52], with a Bland-Altman bias of 957.5 (SD 679.9) kcal, when the mean EE measured with the reference standard was 1456.48 (SD 731.40) kcal [34]. For the *Polar Loop*, the MAPE for EE ranged from 6% to 56% [63]. The

Pearson correlation coefficient was 0.74 ($P < .01$) for the *Polar A300* [16].

For the *Withings Pulse*, the MAPE for EE ranged from -39% to 64% [63,66].

Outcomes Reported Less Frequently

Other outcomes that were evaluated less frequently include distance, reported in 3 studies on 15 devices from 7 brands, always using the measured distance as the reference standard [30,35,63]; speed, reported in a study using 1 device, with actual speed (on a treadmill) as the reference standard [23]; and activity count, defined as the number of activities (eg, number of arm movements or body movements based on observation or measured acceleration data), reported from 4 studies on 4 devices from 4 different brands using as the reference standard manual count (video recording), video analysis (automated), or an activity tracker [17,32,42,54].

Risk of Bias

The risk-of-bias assessment for each outcome is reported in [Multimedia Appendix 1](#). In summary, all the studies were at high or probably high risk of bias for the domain *Patient selection* because they used a convenience sampling technique. Almost all the studies were at low risk of bias for the domains *Index test* and *Reference standard* because the 2 measurement methods were applied at the same time and interpreted without knowledge of the results obtained with the other method. A small number of studies was identified as high risk for the domain *Flow and timing* based on the high percentage (>25%) of missing data for the index test or reference standard.

Acceptability

The acceptability of wrist-wearable activity trackers was assessed in 11 studies on 10 devices from 9 brands.

Data Availability

In all, 36% (4/11) of studies focused on data availability, expressed as a proportion of time in which the data were available, and a different device was used in each of these studies. The denominator for the proportion could be the study duration or the time spent exercising. Rowlands et al [75] found that data availability was 52% in a pediatric healthy population using the *GENEActiv* for 14 months. Deka et al [68] focused on data availability during exercise time. In this study, adult patients with cardiac heart failure activated their *Fitbit Charge HR* in 75% of the exercise sessions (over 5 days) and data were available for 99% of the time when activated. Marcoux et al [72] studied the *Fitbit Flex 2* in adults with idiopathic pulmonary fibrosis (for 46 days). Of the 20 patients, 2 did not succeed in activating the device. Among the remaining participants, data were available for a mean of 91% (SD 20%) of the time. Lahti et al [71] studied the *Garmin Vivofit* in adults with schizophrenia and found data available for 97% of the time (over 4 months).

Wearing Time

In all, 27% (3/11) of studies reported on the wearing time. Farina et al [69], using the *GENEActiv*, found that 89% of the participants with dementia and 86% of their caregivers wore the device for the duration of the study (28 days). Speier et al

[74], using the *Fitbit Charge 2*, enrolled participants with coronary artery disease. The median time spent wearing the activity tracker ranged from 44% to 90% over 90 days. Finally, for *Nike FuelBand*, in a study on patients with schizophrenia, the mean wearing time was 89% (SD 13%) over 80-133 days [73].

Ease of Use and Other Characteristics

In all, 36% (4/11) of studies focused on the ease of use and similar characteristics of wrist-wearing devices. The *Polar A300* was assessed in patients with chronic obstructive pulmonary disease wearing the device for 3 days using the Post-Study System Usability Questionnaire, which calculates a score that ranges from 1 to 7 (the lower the better) for 3 subdomains [16]. The mean scores were 1.46 (SD 0.23) for system quality, 2.41 (SD 0.53) for information quality, and 3.35 (SD 0.62) for interface quality. The *AX3 data logger* was assessed in persons with Parkinson disease wearing the device for 7 days [70]. A questionnaire created ad hoc was used for the assessment; 94% of the participants agreed that they were willing to wear the sensors at home, and 85% agreed that they were willing to wear the sensors in public. However, some of the participants reported problems with the strap fitting and the material (number not reported). The *Fitbit Flex* was assessed with a questionnaire created ad hoc in a study on pregnant women followed for 7 days [57]. The *Fitbit Flex* was reported by 31% to be inconvenient, 6% to be poorly esthetic, and 12% to be uncomfortable. Kaewkannate et al [39] asked healthy participants to wear 4 different devices over 28 days and compared them using a questionnaire created ad hoc. The *Withings Pulse* had the highest satisfaction score, followed by *Misfit Shine*, *Jawbone UP24*, and *Fitbit Flex*.

Discussion

Study Findings

We systematically reviewed the available evidence on the acceptability and accuracy of wrist-wearable activity-tracking devices for measuring physical activity across different devices and measures. We found substantial heterogeneity among the included studies. The main sources of heterogeneity were the studies' population and setting, the device used, the reference standard, the outcome assessed, and the outcome measure reported.

Acceptability was evaluated in 11 studies on 10 devices from 9 brands. Data availability was $\geq 75\%$ for the *Fitbit Charge HR*, *Fitbit Flex 2*, and *Garmin Vivofit*. Data availability is defined as the amount of data captured over a certain time period, which, in this case, is over a predetermined duration of each respective study. Data availability can be a measure of how accurate a device is at capturing data when the device is worn. For example, if an individual wears the device for 8 hours but only 4 hours of data are available, some questions may be raised on the capability of the device to capture information accurately. The wearing time was 89% for both the *GENEActiv* and *Nike FuelBand*. Wearing time is defined as the amount of time the device is worn over a predetermined duration for each study. For each study, wearing time may have been assessed differently; for example, a study may measure wearing time

over a day, whereas another study may measure over a week. Both data availability and wearing time can provide a deeper look into acceptability because participants may wear a device more frequently and, ultimately, have more data available if a device is more acceptable. Accuracy was assessed in 57 studies on 72 devices from 29 brands. Among 14 outcomes assessed, step counts, heart rate, and EE were the ones used most frequently. For step counts, the Fitbit Charge (or the Fitbit Charge HR) had a MAPE <25% across 20 studies. For heart rate, the Apple Watch had a MAPE <10% in 2 studies. For EE, the MAPE was >30% for all the brands, showing poor accuracy across devices.

Comparison With Other Systematic Reviews

Feehan et al [77] conducted a systematic review on the accuracy of Fitbit devices for measuring physical activity. The review did not specifically focus on wrist-wearable activity trackers; it also included studies using activity trackers worn on other body locations (torso, ankle, or hip). This systematic review reported a good accuracy of Fitbit devices in measuring steps, with 46% of the included studies reporting a measurement error within -3% to +3%. Regarding EE, the authors concluded that “Fitbit devices are unlikely to provide accurate measures of energy expenditure.” Studies on heart rate were not included in the review. Evenson et al [78] performed a systematic review focusing on Fitbit and Jawbone devices. Similarly, wearing the device on the wrist was not an inclusion criterion. The authors concluded that for step counts, the included studies often showed a high correlation, with the correlation coefficient ≥ 0.80 among devices from both brands, with the reference standards. The correlation was frequently low for the outcome EE. Similar to the review by Feehan et al [77], the outcome heart rate was not included in this systematic review. The results of these systematic reviews are consistent with our findings for the devices and outcomes assessed.

Strengths and Limitations

The main strengths of our systematic review include the inclusion of all the devices reported in the literature; the reporting on all the outcomes related to acceptability and accuracy, with no restrictions; and the assessment of the risk of bias of the included studies. These characteristics make this review unique for this topic. However, in our systematic review, we decided to exclude studies in which a wearable device was not positioned on a wrist. Some devices can be positioned both on the wrist and other sites (torso, hip, ankle, arm, or brassiere), and the acceptability and accuracy can vary for the same device depending on where it is positioned, increasing heterogeneity [77,78]. Therefore, our results cannot be generalized to the acceptability and accuracy of devices worn on sites other than wrists. Acceptability is defined and measured in many different ways in the literature about wearing devices and about information technology in general [79]. These definitions are often broad and nonspecific, with published literature suggesting that acceptability research should become more robust [80]. For the purpose of our paper, acceptability was operationalized using proxies such as wearing time or data availability. However, other definitions have proposed that acceptability is related more to the extent to which individuals receiving a health

care intervention find it appropriate based on cognitive and emotional responses to the intervention [80]. It is important to recognize that acceptability may be more of a holistic and subjective construct rather than an objective one, and thus wear time or data availability may not do full justice to acceptability. Although these metrics have the advantage of being relatively easy to obtain and reproduce, allowing for quantitative comparisons, they are only proxies for acceptability, which is a more nuanced concept. For example, one might wonder if wearing time is low because a person only wears the device a few hours each day or only on weekends or if they completely stopped wearing it after some time. Moreover, wearing time is more likely to offer valuable information in studies with a long follow-up, whereas 2 out of 3 studies reporting on this outcome had a follow-up of <1 week. Because of the presence of important heterogeneity among studies, we were not able to perform a meta-analysis.

Regardless, the comprehensive reporting in this review will allow researchers to assess the available evidence and inform future studies, either to further assess the accuracy of wearable devices or to inform the choice of one device over another to use in interventional studies. To facilitate these choices, we have provided to readers the database with the results of the individual included studies and we did our best to offer a synthesis of the 3 outcomes reported most frequently (step counts, heart rate, and EE).

Future Research

Further high-quality studies are needed to determine the accuracy and acceptability of wearable devices for measuring physical activity. Given the number of devices available (72 included in this review), it is unlikely that a single study will be able to answer this question. This makes it particularly important to standardize some aspects of these studies, to reduce the heterogeneity among them, and allow for meta-syntheses of the results with comparisons across studies, devices, and outcomes. If the heterogeneity was acceptable, a network meta-analysis would also allow researchers to make indirect comparisons. The main sources of heterogeneity that could be controlled are the setting of the study, the population, the reference standard used, and the outcome definition and measure. A first step in this direction would be putting together a task force of experts to issue guidelines on how to report these experiments, similar to guidelines for the EQUATOR network. A second step would be to issue recommendations on this aspect, starting with accepted reference standards against which devices should be tested for each outcome, the conditions in which the experiment should be conducted, and the way in which the outcomes should be measured and analyzed. Regarding the reference standards, some of these are more accurate than others. Our approach was to take accuracy to mean criterion and convergent validity in this review, but once there is consensus on the acceptable reference standard, other comparisons should not be included in a meta-synthesis. Regarding the method to report on the accuracy of continuous variables (more common in this field), this is the order of priority that we suggest: MAPE, mean percentage error, mean difference, Bland–Altman mean bias, and measure of correlation as the least preferred. This is because the percentage error gives the reader a better

understanding of the importance of the error (a mean error of 50 steps is much more relevant if the total step count was 100 than if it was 10,000). We preferred the MAPE over the mean absolute error because when the absolute value is not used, there is a risk of negative and positive errors balancing each other, with the risk of overestimating the accuracy. We prefer the mean difference over the Bland–Altman mean bias because in an accuracy study, the reference standard is supposed to be more accurate than the index test, and therefore the latter should be tested against the former, not against their mean. In the case of the acceptability outcome, consensus should be reached also on how to define and measure it. For example, defining a minimum set of outcomes to be reported might help in this context. This might include reporting the percentage of abandonment over time. Furthermore, as new devices become available, their acceptability and accuracy should also be tested because they could differ from the acceptability and accuracy of other devices, even those produced by the same company. Regarding the choice of the device to use in interventional studies, for example, in studies that aim at increasing physical activity in a certain population, there is no one-device-fits-all answer. This choice should be based on the available data on

acceptability and accuracy and be tailored to the outcome to measure. In a study with step count as the main outcome, the Fitbit Charge and Fitbit Charge HR might be appropriate choices. The Apple Watch might be preferred if the main outcome is heart rate. Active time was most often measured through time spent in MVPA, and the Fitbit Flex is the only device that was used in 3 studies, showing good results in 2 of these. Regarding EE, we do not feel comfortable suggesting the use of any device based on the current evidence because the accuracy was poor across devices. The decision should probably be driven by the other outcomes used. Broader recommendations should be issued in the form of guidelines from a panel of experts using this systematic review as a knowledge base.

Conclusions

We reported on the acceptability and accuracy of 72 wrist-wearable devices for measuring physical activity produced by 29 companies. The Fitbit Charge and Fitbit Charge HR were consistently shown to have a good accuracy for step counts and the Apple Watch for measuring heart rate. None of the tested devices proved to be accurate in measuring EE. Efforts should be made to reduce the heterogeneity among studies.

Authors' Contributions

FG and AI conceived this review. FG is the guarantor of the review and drafted the manuscript. TN and VBD developed the search strategy. FG, NN, VBD, APB, and DP screened the articles and extracted the data. All the authors read, provided feedback, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Database of result characteristics of the studies reporting on accuracy and acceptability.
[\[XLSX File \(Microsoft Excel File\), 34 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Guide for accessing database in Multimedia Appendix 1.
[\[PDF File \(Adobe PDF File\), 206 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Result characteristics of the studies reporting on accuracy.
[\[DOCX File , 45 KB-Multimedia Appendix 3\]](#)

References

1. Seifert A, Schlomann A, Rietz C, Schelling HR. The use of mobile devices for physical activity tracking in older adults' everyday life. *Digit Health* 2017 Nov 09;3:2055207617740088 [[FREE Full text](#)] [doi: [10.1177/2055207617740088](https://doi.org/10.1177/2055207617740088)] [Medline: [29942617](https://pubmed.ncbi.nlm.nih.gov/29942617/)]
2. Ainsworth BE. How do I measure physical activity in my patients? Questionnaires and objective methods. *Br J Sports Med* 2009 Jan;43(1):6-9. [doi: [10.1136/bjism.2008.052449](https://doi.org/10.1136/bjism.2008.052449)] [Medline: [18718977](https://pubmed.ncbi.nlm.nih.gov/18718977/)]
3. Rütten A, Ziemainz H, Schena F, Stahl T, Stiggelbout M, Auweele YV, et al. Using different physical activity measurements in eight European countries. Results of the European Physical Activity Surveillance System (EUPASS) time series survey. *Public Health Nutr* 2003 Jun;6(4):371-376. [doi: [10.1079/PHN2002450](https://doi.org/10.1079/PHN2002450)] [Medline: [12795825](https://pubmed.ncbi.nlm.nih.gov/12795825/)]
4. Rothwell PM. Analysis of agreement between measurements of continuous variables: general principles and lessons from studies of imaging of carotid stenosis. *J Neurol* 2000 Nov;247(11):825-834. [doi: [10.1007/s004150070068](https://doi.org/10.1007/s004150070068)] [Medline: [11151413](https://pubmed.ncbi.nlm.nih.gov/11151413/)]

5. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. *Perspect Clin Res* 2017;8(4):187-191 [FREE Full text] [doi: [10.4103/picr.PICR_123_17](https://doi.org/10.4103/picr.PICR_123_17)] [Medline: [29109937](https://pubmed.ncbi.nlm.nih.gov/29109937/)]
6. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005 May;37(5):360-363 [FREE Full text] [Medline: [15883903](https://pubmed.ncbi.nlm.nih.gov/15883903/)]
7. Dillon A, Morris M. User acceptance of information technology: theories and models. *Annual Review of Information Science and Technology (ARIST)*. 1996. URL: <https://eric.ed.gov/?id=EJ536186> [accessed 2021-12-30]
8. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. *J Biomed Inform* 2019 May;93:103153. [doi: [10.1016/j.jbi.2019.103153](https://doi.org/10.1016/j.jbi.2019.103153)] [Medline: [30910623](https://pubmed.ncbi.nlm.nih.gov/30910623/)]
9. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016 Dec 05;5(1):210 [FREE Full text] [doi: [10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4)] [Medline: [27919275](https://pubmed.ncbi.nlm.nih.gov/27919275/)]
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529-536. [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
11. 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 Feb 23;1476-1485. [doi: [10.1177/2047487316634883](https://doi.org/10.1177/2047487316634883)] [Medline: [26907794](https://pubmed.ncbi.nlm.nih.gov/26907794/)]
12. 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/)]
13. 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/)]
14. An H, Kim Y, Lee J. Accuracy of inclinometer functions of the activPAL and ActiGraph GT3X+: a focus on physical activity. *Gait Posture* 2017 Jan;51:174-180 [FREE Full text] [doi: [10.1016/j.gaitpost.2016.10.014](https://doi.org/10.1016/j.gaitpost.2016.10.014)] [Medline: [27780084](https://pubmed.ncbi.nlm.nih.gov/27780084/)]
15. Blondeel A, Loeckx M, Rodrigues F, Janssens W, Troosters T, Demeyer H. Wearables to coach physical activity in patients with COPD: validity and patient's experience. *Am J Respirat Crit Care Med* 2018;197:A7072 [FREE Full text]
16. 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/)]
17. Bruder AM, McClelland JA, Shields N, Dodd KJ, Hau R, van de Water AT, et al. Validity and reliability of an activity monitor to quantify arm movements and activity in adults following distal radius fracture. *Disabil Rehabil* 2018 Jun;40(11):1318-1325. [doi: [10.1080/09638288.2017.1288764](https://doi.org/10.1080/09638288.2017.1288764)] [Medline: [28637143](https://pubmed.ncbi.nlm.nih.gov/28637143/)]
18. Bulathsinghala C, Tejada J, Wallack RZ. Validating output from an activity monitor worn on the wrist in patients with COPD. *Chest* 2014 Oct;146(4):25A. [doi: [10.1378/chest.1989979](https://doi.org/10.1378/chest.1989979)]
19. 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/)]
20. Choi L, Chen KY, Acra SA, Buchowski MS. Distributed lag and spline modeling for predicting energy expenditure from accelerometry in youth. *J Appl Physiol* (1985) 2010 Mar;108(2):314-327 [FREE Full text] [doi: [10.1152/jappphysiol.00374.2009](https://doi.org/10.1152/jappphysiol.00374.2009)] [Medline: [19959770](https://pubmed.ncbi.nlm.nih.gov/19959770/)]
21. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. *Gait Posture* 2017 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/)]
22. 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/)]
23. Cohen MD, Cutaia M. A novel approach to measuring activity in chronic obstructive pulmonary disease: using 2 activity monitors to classify daily activity. *J Cardiopulm Rehabil Prev* 2010;30(3):186-194. [doi: [10.1097/HCR.0b013e3181d0c191](https://doi.org/10.1097/HCR.0b013e3181d0c191)] [Medline: [20216326](https://pubmed.ncbi.nlm.nih.gov/20216326/)]
24. Compagnat M, Mandigout S, Chaparro D, Daviet JC, Salle JY. Validity of the Actigraph GT3x and influence of the sensor positioning for the assessment of active energy expenditure during four activities of daily living in stroke subjects. *Clin Rehabil* 2018 Dec;32(12):1696-1704. [doi: [10.1177/0269215518788116](https://doi.org/10.1177/0269215518788116)] [Medline: [30012036](https://pubmed.ncbi.nlm.nih.gov/30012036/)]
25. Dondzila C, Lewis C, Lopez J, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. *Int J Exerc Sci* 2018;11(7):584 [FREE Full text]
26. 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/)]
27. Durkalec-Michalski K, Wozniewicz M, Bajerska J, Jeszka J. Comparison of accuracy of various non-calorimetric methods measuring energy expenditure at different intensities. *Hum Mov* 2013;14(2):167. [doi: [10.2478/humo-2013-0019](https://doi.org/10.2478/humo-2013-0019)]

28. Falgoust B, Handwerger B, Rouly L, Blaha O, Pontiff M, Eason J. Accuracy of wearable activity monitors in the measurement of step count and distance. *Cardiopulm Phys Ther J* 2018;29(1):36 [FREE Full text]
29. 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;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/)]
30. 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/)]
31. Gillinov AM, Etiwy M, Gillinov S, Wang R, Blackburn G, Phelan D, et al. Variable accuracy of commercially available wearable heart rate monitors. *J Am Coll Cardiol* 2017 Mar;69(11):336. [doi: [10.1016/s0735-1097\(17\)33725-7](https://doi.org/10.1016/s0735-1097(17)33725-7)]
32. Girona RJ, Lloyd J, Clark ME, Walker RL. Preliminary evaluation of reliability and criterion validity of Actiwatch-Score. *J Rehabil Res Dev* 2007;44(2):223-230 [FREE Full text] [doi: [10.1682/jrrd.2006.06.0058](https://doi.org/10.1682/jrrd.2006.06.0058)] [Medline: [17551874](https://pubmed.ncbi.nlm.nih.gov/17551874/)]
33. 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/)]
34. 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 [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/)]
35. 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 Dec;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/)]
36. 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* 2017 May 08:844-850. [doi: [10.1136/bjsports-2016-096990](https://doi.org/10.1136/bjsports-2016-096990)] [Medline: [28483930](https://pubmed.ncbi.nlm.nih.gov/28483930/)]
37. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. 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/)]
38. 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. *Int J Sports Phys Ther* 2018 Aug;13(5):860-870 [FREE Full text] [Medline: [30276018](https://pubmed.ncbi.nlm.nih.gov/30276018/)]
39. Kaewkannate K, Kim S. A comparison of wearable fitness devices. *BMC Public Health* 2016 May 24;16:433 [FREE Full text] [doi: [10.1186/s12889-016-3059-0](https://doi.org/10.1186/s12889-016-3059-0)] [Medline: [27220855](https://pubmed.ncbi.nlm.nih.gov/27220855/)]
40. 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/)]
41. Lauritzen J, Muñoz A, Sevillano JL, Civit BA. The usefulness of activity trackers in elderly with reduced mobility: a case study. *Stud Health Technol Inform* 2013;192:759-762. [Medline: [23920659](https://pubmed.ncbi.nlm.nih.gov/23920659/)]
42. Lawinger E, Uhl T, Abel M, Kamineni S. Assessment of accelerometers for measuring upper-extremity physical activity. *J Sport Rehabil* 2015 Aug;24(3):236-243. [doi: [10.1123/jsr.2013-0140](https://doi.org/10.1123/jsr.2013-0140)] [Medline: [25803521](https://pubmed.ncbi.nlm.nih.gov/25803521/)]
43. Lemmens PM, Sartor F, Cox LG, den Boer SV, Westerink JH. Evaluation of an activity monitor for use in pregnancy to help reduce excessive gestational weight gain. *BMC Pregnancy Childbirth* 2018 Jul 31;18(1):312 [FREE Full text] [doi: [10.1186/s12884-018-1941-8](https://doi.org/10.1186/s12884-018-1941-8)] [Medline: [30064390](https://pubmed.ncbi.nlm.nih.gov/30064390/)]
44. Magistro D, Brustio PR, Ivaldi M, Eslinger DW, Zecca M, Rainoldi A, et al. Validation of the ADAMO Care Watch for step counting in older adults. *PLoS One* 2018;13(2):e0190753 [FREE Full text] [doi: [10.1371/journal.pone.0190753](https://doi.org/10.1371/journal.pone.0190753)] [Medline: [29425196](https://pubmed.ncbi.nlm.nih.gov/29425196/)]
45. Mandigout S, Lacroix J, Ferry B, Vuillerme N, Compagnat M, Daviet J. Can energy expenditure be accurately assessed using accelerometry-based wearable motion detectors for physical activity monitoring in post-stroke patients in the subacute phase? *Eur J Prev Cardiol* 2017 Dec;24(18):2009-2016. [doi: [10.1177/2047487317738593](https://doi.org/10.1177/2047487317738593)] [Medline: [29067851](https://pubmed.ncbi.nlm.nih.gov/29067851/)]
46. Manning M, Tune B, Völgyi E, Smith W. Agreement of activity monitors during treadmill walking in teenagers with severe obesity. *Pediatr Exerc Sci* 2016;28:44 [FREE Full text]
47. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. *Meas Phys Educ Exerc Sci* 2017 Jun 08;21(4):201-211. [doi: [10.1080/1091367x.2017.1331166](https://doi.org/10.1080/1091367x.2017.1331166)]
48. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation of a self-monitoring tool for use in exercise therapy. *PM R* 2017 Nov;9(11):1077-1084 [FREE Full text] [doi: [10.1016/j.pmrj.2017.03.012](https://doi.org/10.1016/j.pmrj.2017.03.012)] [Medline: [28400221](https://pubmed.ncbi.nlm.nih.gov/28400221/)]
49. 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/)]
50. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC Med Res Methodol* 2019 Feb 07;19(1):29 [FREE Full text] [doi: [10.1186/s12874-019-0668-1](https://doi.org/10.1186/s12874-019-0668-1)] [Medline: [30732582](https://pubmed.ncbi.nlm.nih.gov/30732582/)]
51. 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/)]

52. 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/)]
53. 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;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/)]
54. Scott JJ, Rowlands AV, Cliff DP, Morgan PJ, Plotnikoff RC, Lubans DR. Comparability and feasibility of wrist- and hip-worn accelerometers in free-living adolescents. *J Sci Med Sport* 2017 Dec;20(12):1101-1106. [doi: [10.1016/j.jsams.2017.04.017](https://doi.org/10.1016/j.jsams.2017.04.017)] [Medline: [28501418](https://pubmed.ncbi.nlm.nih.gov/28501418/)]
55. Semanik P, Lee J, Pellegrini CA, Song J, Dunlop DD, Chang RW. Comparison of physical activity measures derived from the Fitbit Flex and the ActiGraph GT3X+ in an employee population with chronic knee symptoms. *ACR Open Rheumatol* 2020 Jan 02;2(1):48-52 [FREE Full text] [doi: [10.1002/acr2.11099](https://doi.org/10.1002/acr2.11099)] [Medline: [31943970](https://pubmed.ncbi.nlm.nih.gov/31943970/)]
56. Sirard JR, Masteller B, Freedson PS, Mendoza A, Hickey A. Youth oriented activity trackers: comprehensive laboratory- and field-based validation. *J Med Internet Res* 2017 Jul 19;19(7):e250 [FREE Full text] [doi: [10.2196/jmir.6360](https://doi.org/10.2196/jmir.6360)] [Medline: [28724509](https://pubmed.ncbi.nlm.nih.gov/28724509/)]
57. 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/)]
58. Stackpool CM. Accuracy of various activity trackers in estimating steps taken and energy expenditure. UW-L Master's Theses. 2013. URL: <https://minds.wisconsin.edu/handle/1793/66292> [accessed 2021-12-31]
59. Stiles VH, Griew PJ, Rowlands AV. Use of accelerometry to classify activity beneficial to bone in premenopausal women. *Med Sci Sports Exerc* 2013 Dec;45(12):2353-2361. [doi: [10.1249/MSS.0b013e31829ba765](https://doi.org/10.1249/MSS.0b013e31829ba765)] [Medline: [23698245](https://pubmed.ncbi.nlm.nih.gov/23698245/)]
60. 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;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/)]
61. 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/)]
62. 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/)]
63. 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 Sep;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/)]
64. 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/)]
65. Wang L, Liu T, Wang Y, Li Q, Yi J, Inoue Y. Evaluation on step counting performance of wristband activity monitors in daily living environment. *IEEE Access* 2017;5:13020-13027. [doi: [10.1109/access.2017.2721098](https://doi.org/10.1109/access.2017.2721098)]
66. Woodman JA, Crouter SE, Bassett DR, Fitzhugh EC, Boyer WR. 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/)]
67. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENE: optimum sampling frequency and number of axes. *Med Sci Sports Exerc* 2012 Nov;44(11):2228-2234. [doi: [10.1249/MSS.0b013e31825e19fd](https://doi.org/10.1249/MSS.0b013e31825e19fd)] [Medline: [22617400](https://pubmed.ncbi.nlm.nih.gov/22617400/)]
68. Deka P, Pozehl B, Norman JF, Khazanchi D. Feasibility of using the Fitbit Charge HR in validating self-reported exercise diaries in a community setting in patients with heart failure. *Eur J Cardiovasc Nurs* 2018 Oct;17(7):605-611. [doi: [10.1177/1474515118766037](https://doi.org/10.1177/1474515118766037)] [Medline: [29546995](https://pubmed.ncbi.nlm.nih.gov/29546995/)]
69. Farina N, Sherlock G, Thomas S, Lowry RG, Banerjee S. Acceptability and feasibility of wearing activity monitors in community-dwelling older adults with dementia. *Int J Geriatr Psychiatry* 2019 Apr;34(4):617-624. [doi: [10.1002/gps.5064](https://doi.org/10.1002/gps.5064)] [Medline: [30701592](https://pubmed.ncbi.nlm.nih.gov/30701592/)]
70. Fisher JM, Hammerla NY, Rochester L, Andras P, Walker RW. Body-worn sensors in Parkinson's disease: evaluating their acceptability to patients. *Telemed J E Health* 2016 Jan;22(1):63-69 [FREE Full text] [doi: [10.1089/tmj.2015.0026](https://doi.org/10.1089/tmj.2015.0026)] [Medline: [26186307](https://pubmed.ncbi.nlm.nih.gov/26186307/)]
71. Lahti A, White D, Katiyar N, Pei H, Baker S. 325. Clinical utility study towards the use of continuous wearable sensors and patient reported surveys for relapse prediction in patients at high risk of relapse in schizophrenia. *Biolog Psychiatry* 2017 May;81(10):S133. [doi: [10.1016/j.biopsych.2017.02.340](https://doi.org/10.1016/j.biopsych.2017.02.340)]
72. Marcoux V, Fell C, Johansson K. Daily activity trackers and home spirometry in patients with idiopathic pulmonary fibrosis. *Am J Respirat Crit Care Med* 2018;197:A1586 [FREE Full text]
73. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. *Telemed J E Health* 2015 Mar;21(3):213-216. [doi: [10.1089/tmj.2014.0105](https://doi.org/10.1089/tmj.2014.0105)] [Medline: [25536190](https://pubmed.ncbi.nlm.nih.gov/25536190/)]

74. Speier W, Dzibur E, Zide M, Shufelt C, Joung S, Van Eyk JE, et al. Evaluating utility and compliance in a patient-based eHealth study using continuous-time heart rate and activity trackers. *J Am Med Inform Assoc* 2018 Oct 01;25(10):1386-1391. [doi: [10.1093/jamia/ocy067](https://doi.org/10.1093/jamia/ocy067)] [Medline: [29850807](https://pubmed.ncbi.nlm.nih.gov/29850807/)]
75. Rowlands AV, Harrington DM, Bodicoat DH, Davies MJ, Sherar LB, Gorely T, et al. Compliance of adolescent girls to repeated deployments of wrist-worn accelerometers. *Med Sci Sports Exerc* 2018 Jul;50(7):1508-1517. [doi: [10.1249/MSS.0000000000001588](https://doi.org/10.1249/MSS.0000000000001588)] [Medline: [29474208](https://pubmed.ncbi.nlm.nih.gov/29474208/)]
76. Noun Project: Free Icons & Stock Photos for Everything. URL: <https://thenounproject.com/> [accessed 2021-09-11]
77. 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/)]
78. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;12(1):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/)]
79. Nadal C, Doherty G, Sas C. Technology acceptability, acceptance and adoption-definitions and measurement. In: Proceedings of the CHI 2019 Conference on Human Factors in Computing Systems. 2019 Presented at: CHI 2019 Conference on Human Factors in Computing Systems; May 4–9, 2019; Glasgow, Scotland UK URL: https://www.researchgate.net/publication/331787774_Technology_acceptability_acceptance_and_adoption_-_definitions_and_measurement
80. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Serv Res* 2017 Jan 26;17(1):1-13 [FREE Full text] [doi: [10.1186/s12913-017-2031-8](https://doi.org/10.1186/s12913-017-2031-8)] [Medline: [28126032](https://pubmed.ncbi.nlm.nih.gov/28126032/)]

Abbreviations

EE: energy expenditure

MAPE: mean absolute percentage error

MVPA: moderate to vigorous-intensity physical activity

Edited by A Mavragani; submitted 28.05.21; peer-reviewed by F Yu, M Western; comments to author 30.07.21; revised version received 24.09.21; accepted 06.12.21; published 21.01.22

Please cite as:

Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, Keepanasseril A, Parpia S, de Wit K, Iorio A Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature

J Med Internet Res 2022;24(1):e30791

URL: <https://www.jmir.org/2022/1/e30791>

doi: [10.2196/30791](https://doi.org/10.2196/30791)

PMID:

©Federico Germini, Noella Noronha, Victoria Borg Debono, Binu Abraham Philip, Drashti Pete, Tamara Navarro, Arun Keepanasseril, Sameer Parpia, Kerstin de Wit, Alfonso Iorio. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 21.01.2022. 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 the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.