HHS Public Access
Author manuscript
J Meas Phys Behav. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:
J Meas Phys Behav. 2020 June ; 3(2): 170–185. doi:10.1123/jmpb.2019-0035.

Review of Validity and Reliability of Garmin Activity Trackers

Kelly R. Evenson, PhD, MS,
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, United States

Camden L. Spade
Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, United States

Abstract

Purpose: A systematic review to summarize the validity and reliability of steps, distance, energy expenditure, speed, elevation, heart rate, and sleep assessed by Garmin activity trackers.

Methods: Searches included studies published through December 31, 2018. Correlation coefficients (CC) were assessed as low (<0.60), moderate (0.60-<0.75), good (0.75-<0.90), or excellent (>=0.90). Mean absolute percentage errors (MAPE) were assessed as acceptable at <5% in controlled conditions and <10% for free-living.

Results: Overall, 32 studies of adults documented validity. Four of these studies also documented reliability. The sample size ranged from 1 to 95 for validity and 4 to 31 for reliability testing. Step inter- and intra-reliability was good-to-excellent and speed intra-reliability was excellent. No other features were explored for reliability. Step validity, across 16 studies, generally indicated good-to-excellent CC and acceptable MAPE. Distance validity, tested in three studies, generally indicated poor CC and MAPE that exceeded acceptable limits, with both over and underestimation. Energy expenditure validity, across 12 studies, generally indicated wide variability in CC and MAPE that exceeded acceptable limits. Heart rate validity in five studies had low-to-excellent CC and all MAPE exceeded acceptable limits. Speed, elevation, and sleep validity were assessed in only one or two studies each; for sleep, the criterion relied on self-report rather than polysomnography.

Conclusion: This systematic review of Garmin activity trackers among adults indicated higher validity of steps; few studies on speed, elevation, and sleep; and lower validity for distance, energy expenditure, and heart rate. Intra- and inter-device feature reliability needs further testing.

Keywords

activity tracker; digital health; heart rate; physical activity; steps; wearables

Address for Correspondence: Kelly R. Evenson, 123 W Franklin Street, Building C, Suite 410, University of NC, Gillings School of Global Public Health, Department of Epidemiology, Chapel Hill, North Carolina, USA 27516, kelly_evenson@unc.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

Wearables are worn devices that can provide a variety of feedback. From a search conducted in 2017, 423 unique wearables distributed across 132 brands were identified (Henriksen et al., 2018). This was an increase from only 3 wearables identified in 2011. In line with the proliferation of wearables, based on a 2018 survey of more than two thousand health professionals from around the world, "wearable technology" was considered the leading fitness trend (Thompson, 2019).

Activity trackers, a subset of wearables, have quickly caught on for personal use, such as to promote changes in physical activity (Strath and Rowley, 2018). In support of this, the Community Guide recommended activity trackers to increase physical activity among overweight or obese adults (de Vries et al., 2016). Consumers are also using activity trackers to communicate with healthcare providers and make more informed health-related decisions (Strath and Rowley, 2018; Wright et al., 2017). In addition, activity trackers are being extensively used for research purposes, both for intervention and measurement, as indicated in both the clinicaltrials.gov database of clinical trials and in the National Institutes of Health RePORTER database of United States' governmental funded studies (Wright et al., 2017). Researchers who wish to use activity trackers must decide from a plethora of device options and features.

With the rise in the choice of activity trackers comes the integration of new sensors that can provide diverse features to the devices, including photoplethysmography, global positioning systems (GPS), barometry, and altimetry (Henriksen et al., 2018). When researchers consider which activity tracker to use, best practice indicates that the information output from the device (i.e., features) should be both valid and reliable (Duking et al., 2018). However, the literature assessing activity trackers is voluminous, with varied protocols, brands and versions, locations worn, and modes of testing. This makes it challenging to assess which device and features within devices to use for research purposes.

Systematic reviews on activity trackers from the same company offer the opportunity to document the history and lineage of their devices. Activity trackers are probably operationally more similar within company than across companies. For example, proprietary algorithms differ across companies and are likely repurposed within the same company. Previously, this type of review was conducted for Fitbit and Jawbone devices (Evenson et al., 2015). We proposed a similar review on Garmin activity trackers.

Garmin (Garmin Ltd., Olathe, Kansas) was founded in 1989 and, as early as 2006 offered activity trackers. Based on second quarter 2018, Garmin ranked fifth in amount of shipments worldwide of activity trackers at 5.3% (International Data Corporation, 2018). In December 2018, an announcement indicated that Garmin would be partnering with ActiGraph, one of the leaders in research-grade accelerometry, for a future product (Muoio, 2018; Plasqui et al., 2013). Garmin devices are also being used in clinical settings both for intervention and measurement. Conducting a search in the clinical trials database (clinicaltrials.gov) on December 16, 2019 revealed 41 studies using a Garmin wearable device.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In order to facilitate use of activity trackers in research, we conducted a systematic review of Garmin activity trackers. Specifically, we summarized the validity and reliability of wrist-worn Garmin activity trackers to assess steps, distance, energy expenditure, speed, elevation, heart rate, and sleep.

Methods

Literature Search

Searches of PubMed, Web of Science, and SPORTDiscus were conducted to include only full-length studies through December 31, 2018. The final search is described in Appendix 1. No start date was imposed in the search. The studies identified from the searches were compiled into Covidence (Melbourne, Victoria) and the two authors selected abstracts for full text review.

Abstracts, conference proceedings, and papers that did not provide the full text in English were excluded. Validity and reliability studies of Garmin trackers that were not activity trackers (example Duncan et al., (2007) were excluded. Studies focused on special populations that might have gait or mobility impairments which could impact the measures under study (examples: Lamont et al., (2018) Madigan, (2019) or Treacy et al., (2017) were also excluded. The review focused on locomotor speed and distance; therefore, we did not include other measures of speed and distance, such as assessed through skiing (Gloersen et al., 2018) or swimming (Mooney et al., 2017). The review also focused on heart rate measured at the wrist; assessment of heart rate straps worn in conjunction with the Garmin wrist-worn activity tracker were not included (for example Cassirame et al., (2017).

Abstraction and Analysis

First, descriptive information on the activity trackers (models, release date, placement, size, weight, and cost) from the Garmin website was recorded. Second, an abstraction tool used for this review was expanded from a tool initially created by De Vries et al. (2009) to document study characteristics and measurement properties of the activity trackers. Specifically, we extracted information on the study population, protocol, statistical analysis, and results related to validity and reliability. A primary reviewer extracted details and a second reviewer checked each entry, with discrepancies resolved by consensus. For abstracted information missing from the publication, we attempted to contact at least one study author to obtain the information. In total, we contacted authors from 15 papers, among which 12 responded. Summary tables were created from the abstracted information.

Reliability of the activity trackers included (Duking et al., 2018): (i) *intra-device reliability*: defined as reproducibility within the same tracker; and (ii) *inter-device reliability*: defined as reproducibility with different trackers. Validity of the activity trackers included (Higgins and Straub, 2006) (i) c*riterion validity*, defined by comparing the trackers to a criterion measure; and (ii) *construct validity*, defined by comparing the trackers to other constructs that should track or correlate positively (*convergent validity*) or negatively (*divergent validity*).

If reported, we abstracted correlation coefficients (CC). We interpreted the CC using the following ratings: <0.60 low, 0.60-<0.75 moderate, 0.75-<0.90 good, and >=0.90 excellent.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

If reported, we abstracted the mean percentage error (MPE) which captured over- and underestimation, defined as the [(criterion value minus Garmin tracker value)/criterion value]*100. If reported, we also abstracted the mean absolute percentage error (MAPE) which captured the magnitude of mis-estimation, defined as the absolute value of [(criterion value minus Garmin tracker value)/criterion value]*100. The smaller MAPE represented better accuracy and accounted for both over- and underestimation. We interpreted a MAPE<5% in laboratory or controlled conditions (Fokkema et al., 2017) and MAPE<10% in free-living conditions (Chen et al., 2016; Crouter et al., 2003; Nelson et al., 2016; Tudor-Locke et al., 2006) as significantly equivalent to the criterion measure. Anything over those measures was considered a practically relevant difference. We also summarized results from the Bland-Altman plots when presented (Bland and Altman, 1986).

Reporting study quality is standard practice for systematic reviews. However, we could locate no assessment tools specific to testing validity and reliability of a device. Therefore, we developed a 10-item assessment, guided both by a paper describing reporting suggestions for wearable sensors (Duking et al., 2018) and a critical appraisal tool developed originally to assess the quality of cross-sectional studies (Downes et al., 2016). The questions asked:

1. Was the research questions clearly stated?
2. Was the study population clearly defined?
3. Was the testing protocol clearly specified?
4. Is the way the tracker is worn on the wrist specified? (e.g., dominant or non-dominate hand, randomized)
5. Were free-living activities included in the protocol?
6. Were usability results presented?
7. Were the app set-up details described for the Garmin activity tracker?
8. Was the threat for specification error (gold standard not used) minimized?
9. Was intra-device reliability included?
10. Was inter-device reliability included?

Yes or no responses were recorded for all 10 items, with “yes” indicating higher study quality.

Results

In total, the search captured 164 unique papers (including 3 papers identified using other sources), with 42 receiving full text review and 32 studies included in the review (Appendix 2). All 32 studies documented validity and 4 of these studies also documented reliability of Garmin activity trackers. Trackers assessed for validity included the Forerunner 225, 235, 305, 310XT, 910XT, and 920XT; Vivoactive; Vivofit, Vivofit 2; and Vivosmart, Vivosmart HR, and Vivosmart HR+ (Table 1). Trackers assessed for reliability included the Forerunner 305, Vivofit, and Vivosmart. All of these products were wrist-worn, with detailed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

descriptions found in Appendix 3. Although the search was not limited by age, all studies enrolled adults only.

Studies were conducted in Australia (n=1), Belgium (n=1), Canada (n=3), China (n=2), Czech Republic (n=1), Denmark (n=2), Egypt (n=1), Germany (n=1), Ireland (n=1), Italy (n=1), the Netherlands (n=1), Switzerland (n=4), Taiwan (n=1), and the United States (n=13) (Table 2). One study reported two countries (Canada and United States) (Reddy et al., 2018). Data collection study dates ranged from 2014 to 2018, as well as one study in 2012 (Menaspà et al., 2014).

The sample size ranged from 1 (Menaspà et al., 2014) to 95 (Brooke et al., 2017) for validity and 4 (O'Connell et al., 2016) to 31 (Fokkema et al., 2017) for reliability testing. The mean percentage of female participants ranged from 0 (Ammann et al., 2016) to 80 (Hochsmann et al., 2018). The assessment of steps, distance, speed, elevation, energy expenditure, heart rate, and sleep is summarized next, with reliability presented first followed by validity evidence. Study quality, along with the questions used for the assessment, is reported in Appendix 4 for each study.

Steps

A assessment of inter-device reliability of steps from 30 Vivofits indicated very small mean differences while on the treadmill (0 to 5 step mean difference over 5 minutes at each of four speeds), but larger differences when compared to carrying a bag (16 step mean difference over 5 minutes) or pushing a stroller (37 step mean difference over 5 minutes) (Appendix 5) (Chen et al., 2016). Another assessment of inter-device reliability of steps from 4 Vivofits indicated a 13.7% difference between units (O'Connell et al., 2016). An assessment of intra-device reliability (n=30–31), comparing steps from the same Vivosmart at two different treadmill sessions, indicated an acceptable MAPE (1.2–3.5%) during three treadmill speeds, but a larger variation in ICC's (0.51 to 0.79) (Fokkema et al., 2017).

Sixteen studies assessed validity of the Garmin activity trackers to assess steps including the: Forerunner 920XT (Wahl et al., 2017), Vivoactive (Wahl et al., 2017), Vivofit (Alsubheen et al., 2016; An et al., 2017; Chen et al., 2016; Ehrler et al., 2016; El-Amrawy and Nounou, 2015; Huang et al., 2016; O'Connell et al., 2016; Simunek et al., 2016; Wahl et al., 2017), Vivofit 2 (Gaz et al., 2018; Hochsmann et al., 2018; Leth et al., 2017; Munck et al., 2018; Wang et al., 2017), and Vivosmart with (Sears et al., 2017) and without heart rate (Fokkema et al., 2017; Wahl et al., 2017) (Appendix 6). Assessments occurred mostly in the laboratory, although some studies included field-based testing or at-home monitoring (An et al., 2017; Gaz et al., 2018; Huang et al., 2016; Simunek et al., 2016; Wahl et al., 2017; Wang et al., 2017). Criterion measured steps were compared against video observation (Alsubheen et al., 2016; Chen et al., 2016; Ehrler et al., 2016; Hochsmann et al., 2018; Huang et al., 2016; O'Connell et al., 2016; Wahl et al., 2017), gait measurement and analysis device (Wahl et al., 2017), hand-tally of steps (An et al., 2017; El-Amrawy and Nounou, 2015; Fokkema et al., 2017; Gaz et al., 2018; Munck et al., 2018; Sears et al., 2017), a pedometer (An et al., 2017; Simunek et al., 2016), and an accelerometer (Leth et al., 2017; Simunek et al., 2016; Wang et al., 2017).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Generally the activity trackers underestimated steps taken on the treadmill (Alsubheen et al., 2016; Chen et al., 2016; Gaz et al., 2018; Hochsmann et al., 2018), except while on an incline (Alsubheen et al., 2016). Agreement, as indicated by CC between the Garmin activity trackers and walking or running on the treadmill, was good to excellent for the Forerunner 920XT (Wahl et al., 2017), Vivosmart (Fokkema et al., 2017; Wahl et al., 2017), Vivofit (Wahl et al., 2017), and Vivoactive (Wahl et al., 2017), but lower for the Vivosmart HR at 3.5 and 4.0 mph (Sears et al., 2017) (Figure 1a). The CC were lower with faster speed only for the Vivofit (Wahl et al., 2017) and Vivosmart HR (Sears et al., 2017).

MAPE was acceptable (<5%) at treadmill speeds 2 to 3 mph across activity trackers (An et al., 2017; Fokkema et al., 2017; Hochsmann et al., 2018; Wahl et al., 2017; Wang et al., 2017) (Figure 1b). Between 3.1 to 4.0 mph, the MAPE exceeded 5% in several studies (An et al., 2017; Fokkema et al., 2017), but not in others (Chen et al., 2016; Hochsmann et al., 2018). Between 4.1 to 8.1 mph, the MAPE never exceeded 5% (An et al., 2017; Chen et al., 2016; Wahl et al., 2017). However, other studies found higher error with slower walking speeds (Ehrler et al., 2016; Munck et al., 2018).

Other studies explored validity of the activity trackers to assess steps beyond the treadmill. The Vivofit underestimated steps when walking on flat ground and upstairs, but overestimated walking downstairs (Huang et al., 2016). Two other validation studies reported excellent agreement for above-ground walking for the Vivofit (El-Amrawy and Nounou, 2015) and Vivofit 2 (Leth et al., 2017). For the Vivofit, MAPE was acceptable (<5%) for slower but not faster speeds on the track (An et al., 2017), while another study found acceptable MAPE across a variety of activities except when pushing a stroller (Chen et al., 2016). One study tested various surfaces and found that steps on the Vivofit varied slightly across surfaces (e.g., natural lawn, gravel, linoleum, asphalt, ceramic tile) but the MAPE remained acceptable (O’Connell et al., 2016). In a study wherein participants wore the Vivofit at home, MAPE was large (17.8%), but the Pearson CC to another device (New Lifestyles pedometer) was excellent (An et al., 2017). In another study where the Vivofit and Yamax pedometer were worn for one week, the Vivofit underestimated daily steps (Simunek et al., 2016).

Distance

No studies reporting on reliability of Garmin-measured distance were identified. Three studies assessed validity of the Garmin activity trackers to assess distance including the Forerunner 920XT (Wahl et al., 2017), the Vivoactive (Wahl et al., 2017), the Vivofit (Huang et al., 2016; Wahl et al., 2017), the Vivofit 2 (Gaz et al., 2018), and the Vivosmart (Wahl et al., 2017) (Appendix 6). Criterion assessments included both known treadmill distance (Gaz et al., 2018; Huang et al., 2016; Wahl et al., 2017) and measured outdoor distance (Gaz et al., 2018; Huang et al., 2016; Wahl et al., 2017).

Generally, the CC for assessing distance were poor (Figure 2). Starting with the most comprehensive study that included four Garmin activity trackers, distance was overestimated at slower treadmill speeds and underestimated at faster treadmill speeds (Wahl et al., 2017). Another study indicated that the Vivofit overestimated distance during level walking, with the MPE highest at slower walking speeds, and greatly overestimated distance when

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

traveling both up and down stairs (Huang et al., 2016). Another study concurred with the overestimation of distance at slower treadmill speeds, but an underestimation while walking on their own (Gaz et al., 2018).

Energy Expenditure

No studies reporting on reliability of Garmin-measured energy expenditure were identified. Twelve studies assessed validity of the Garmin activity trackers to assess energy expenditure including the: Forerunner 225 (Dooley et al., 2017), Forerunner 305 (Hongu et al., 2013), Forerunner 920XT (Roos et al., 2017; Wahl et al., 2017), Vivoactive (Wahl et al., 2017), Vivofit (Alsubheen et al., 2016; Brooke et al., 2017; Pribyslavska et al., 2018; Price et al., 2017; Wahl et al., 2017; Woodman et al., 2017), Vivofit 2 with a chest strap (Yavelberg et al., 2018), and Vivosmart with (Boudreaux et al., 2018; Reddy et al., 2018) and without heart rate (Wahl et al., 2017) (Appendix 7).

Generally, CC comparing agreement ranged from low to substantial (Boudreaux et al., 2018; Brooke et al., 2017; Price et al., 2017; Reddy et al., 2018; Wahl et al., 2017), with high variability across devices and studies (Figure 3a). In most cases, the MAPE was unacceptable (Figure 3b) (Boudreaux et al., 2018; Brooke et al., 2017; Dooley et al., 2017; Pribyslavska et al., 2018; Reddy et al., 2018; Roos et al., 2017; Wahl et al., 2017; Woodman et al., 2017). The MPE was also large for many different activities (Pribyslavska et al., 2018; Reddy et al., 2018). Three studies not reporting CC or MAPE found large mean differences between the Garmin assessment of energy expenditure and the criterion measure during physical activity (Alsubheen et al., 2016; Hongu et al., 2013; Yavelberg et al., 2018).

Speed

An assessment of intra-device reliability of speed from the Forerunner 305 indicated good to excellent agreement, with ICC's ranging from 0.84 to 0.99 while running at different conditions on a track (Appendix 5) (Hovsepian et al., 2014). This was also the only study to report validity of speed measurement compared to recordings on a track using photoelectric timing lights. For 13 participants, generally the Forerunner slightly underestimated speed (Appendix 8), with the agreement ranging from good to excellent.

Elevation

No studies reporting on reliability of Garmin-measured elevation were identified. Two studies assessed validity to assess elevation using the Forerunner 310XT (Menaspà et al., 2014) and Forerunner 910XT (Ammann et al., 2016) (Appendix 8). In the earlier study, a Forerunner and two SRM PowerControl 7 devices mounted to a car roof rack were compared over 6 tests, repeating the same 16 kilometer mountain climb at different times of day and weather conditions (Menaspà et al., 2014). The Forerunner over estimated elevation, with smaller differences found when elevation correction was not used. The latter study conducted 40 trials for 3 participants using four speeds on a level track, with any elevation gained assumed to be error (Ammann et al., 2016). Across the four speeds, the hip recording (secured by using the wrist strap mounted to a waist-worn belt) produced less elevation gained compared to the wrist recording. At the wrist, where 15% of recordings were outliers, error was higher as speed increased.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Heart Rate

No studies reporting on reliability of Garmin-measured heart rate were identified. Five studies reported on validity using the Forerunner 225 (Claes et al., 2017; Dooley et al., 2017), Forerunner 235 (Gillinov et al., 2017), and the Vivosmart HR+ (Boudreaux et al., 2018) (Appendix 8). Two studies used a Polar chest transmitter to assess heart rate as the criterion measure (Dooley et al., 2017; Reddy et al., 2018), while three studies used a 3- to 12-lead electrocardiogram (ECG) (Boudreaux et al., 2018; Claes et al., 2017; Gillinov et al., 2017).

Three studies assessed the Forerunner tracker, with CC lower for activities that used arms (e.g., elliptical), but higher for rest and treadmill locomotion on flat or elevated grades (Claes et al., 2017; Gillinov et al., 2017) (Figure 4a). However, all MAPE exceeded 5% across rest and various laboratory activities (Figure 4b) (Dooley et al., 2017; Gillinov et al., 2017). For example, 25 participants in a laboratory-based study assessed heart rate using the Forerunner 235 compared to a 12-lead ECG (Gillinov et al., 2017). The MAPE was 6% at rest, and was higher with increasing intensity, particularly when arm movement was involved. Based on the Bland Altman plots, heart rate varied widely across the range of intensity, with 95% of the values falling between −27 to 33 beats/minute of the ECG value.

Two studies assessed heart rate recordings using the Vivosmart, with CC varying widely across activities and the MAPE exceeding 5% in all cases (Figure 4) (Boudreaux et al., 2018; Reddy et al., 2018), with the MPE and Bland Altman plots indicating generally an underestimate of heart rate (Reddy et al., 2018). In one study (Reddy et al., 2018), heart rate assessment was best when the activity mode setting was used. In addition, this study assessed the Vivosmart HR+ while off the body, simulating motion on a shaker table, and found spurious heart rate recordings. In the second study comparing to ECG recorded heart rate, the Vivosmart heart rate values differed from the ECG heart rate values for 10 of the 12 resistance exercises, underestimating heart rate during all 12 of them (Boudreaux et al., 2018).

Sleep

No studies reporting on reliability of Garmin-measured sleep were identified. Two studies assessed validity using the Vivofit (Brooke et al., 2017) and the Vivosmart (Lee et al., 2018) (Appendix 8). The earlier study included 24 participants who wore the Vivofit for two days, enabled sleep mode at bedtime, and kept a sleep log as the criterion measure (Brooke et al., 2017). Mean sleep time was similar between measures, with good CC and acceptable MAPE. The latter study included 40 participants who wore the Vivosmart (Lee et al., 2018). Mean sleep time was overestimated, with low agreement compared to diary measures. Other measures of sleep (e.g., time in bed, sleep efficiency, wake after sleep onset) were also not well measured.

Discussion

This review summarized the evidence for validity and reliability of Garmin activity trackers, identifying 32 studies published between 2013 to 2018. Specifically, the features of steps,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

distance, energy expenditure, speed, elevation, heart rate, and sleep were reviewed, with limited studies on reliability and variation for validity findings. All studies enrolled adults only.

Steps

During controlled testing in the laboratory, in most cases the Garmin activity trackers assessed steps appropriately. However, there were studies indicating exceptions to this between 3.1 to 4.0 mph (An et al., 2017; Fokkema et al., 2017; Huang et al., 2016). Moreover, one study indicated the Vivosmart HR step counts were not correlated with hand counted step counts at faster treadmill speeds (Sears et al., 2017). The tendency was for the Garmin to underestimate steps on a treadmill at 0% grade (Alsubheen et al., 2016; Chen et al., 2016; Gaz et al., 2018; Hochsmann et al., 2018); this trend did not follow while on an incline (Alsubheen et al., 2016) or walking upstairs (Huang et al., 2016). In uncontrolled settings, the performance was similar to previously validated pedometers, with steps both over- and underestimated compared to the criterion (An et al., 2017; Simunek et al., 2016). One study indicated that arm movements seemed to exacerbate error (Chen et al., 2016).

Three studies assessed reliability of step measures, the most of any other feature. Findings indicated that an improvement in intra- and inter-device reliability could help contribute to more stable validity results. The adequate performance of Garmin activity trackers to count steps is in line with reviews of Fitbit and Jawbone activity trackers (Evenson et al., 2015) and with a review of a variety of activity trackers worn by older adults (Straiton et al., 2018). Step performance can be improved by setting the participant's stride length if possible.

Distance

Distance was not well measured using the Garmin activity trackers. Most trials found the trackers over-estimated at slower speeds, including when walking up or down stairs, and under-estimated at faster speeds (Gaz et al., 2018; Huang et al., 2016; Wahl et al., 2017). Other brands of activity trackers also over-estimate distance at slower speeds and under-estimate at faster speeds (Evenson et al., 2015). Stair walking may be particularly problematic since stride length differences can vary. We hypothesize that Garmin activity trackers that use GPS and altimeters to assess distance should be more accurate (Gaz et al., 2018), none of which have been explored by studies through the year 2018.

Energy Expenditure

Generally the amount of error was substantial when comparing the Garmin activity tracker assessment of energy expenditure to a criterion measure. This finding is in line with the validity of energy expenditure assessment from other activity trackers as well (Evenson et al., 2015). The devices tested may only be able to detect gross increases in energy expenditure as reflected in exercise intensity. Both over- and under-estimation of kilocalories were detected. Garmin documentation indicates that "resting calories" or resting metabolic rate is based on age, gender, height, and weight (Garmin, 2019a). "Active calories" is additionally based on activity level, type of activity, and heart rate (if available). Together the resting and active calories sum to total calories. Therefore, user-defined age, gender, height, and weight can impact the estimate, as well as the accuracy of the accelerometer and heart

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rate estimation. Given that heart rate assessment was generally poor, this might weaken the validity of energy expenditure.

Speed

Only one study assessed the reliability and validity of the assessment of speed (Hovsepian et al., 2014). Using a Forerunner, both reliability and validity ranged from good to excellent on a track surface. This study also tested a second wearable device (Polar RS800cx with footpod), and found somewhat higher reliability and validity in the same test conditions as compared to the Forerunner. Conclusions are challenging for speed, given that only one study was identified.

Elevation

While the reliability of elevation is not known, elevation was over estimated in two validation studies (Menaspà et al., 2014). The two studies tested the trackers in different scenarios. The first study compared elevation gained up a mountain climb using two SRM PowerControl devices which utilized a barometric altimeter to determine elevation (Menaspà et al., 2014). Since Garmin activity trackers assessed position in a horizontal plane reliably, cross-referencing elevation based on GPS position to elevation data from professional surveys should improve the reliability of elevation measurement (Menaspà et al., 2014). However, this small study indicated that elevation correction exacerbated rather than address the problem. The second study performed testing on a level track and assessed any elevation gained as error (Ammann et al., 2016). Error increased as speed increased, with more error found at the wrist than the hip. The authors attributed the arm swing in overestimating elevation gained, and recommended hip placement for more accurate assessment. These two studies highlighted how elevation measurement can be altered, and further assessment across a range of devices is needed.

Heart Rate

The three studies (Claes et al., 2017; Dooley et al., 2017; Gillinov et al., 2017) of the Forerunner indicated that heart rate assessment was better at rest than with physical activity, and degraded when arm movements were involved in the activity. In most cases, the measures exceeded the 5% MAPE that we used as an acceptable level for laboratory assessments. These studies reported both under- and over-estimation of heart rate. For the two studies of the Vivosmart (Boudreaux et al., 2018; Reddy et al., 2018), similar findings emerged, with a wide range in agreement between the two heart rate measures, and degradation of concordance with increasing physical activity intensity. Validity studies that assessed heart rate using other brands of activity trackers found similar results (Boudreaux et al., 2018; Cadmus-Bertram et al., 2017; Dooley et al., 2017; Gillinov et al., 2017; Jo et al., 2016; Wallen et al., 2016), so the challenges are not inherent only to Garmin activity trackers. When comparing both wrist-worn activity trackers and Polar-worn chest straps to ECG, heart rate was more accurate using the chest strap (Gillinov et al., 2017). Garmin offers chest straps for some devices; it is logical to hypothesize that this would improve the accuracy of the heart rate reading.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Garmin wrist-worn trackers that assess heart rate without a chest strap use optical light sensors called photoplethysmography (Garmin, 2019b). Heart rate is based on the differential reflection of these light emitting diodes in response to the pulsatile changes in blood volume with each heart contraction near the skin surface (Reddy et al., 2018). According to Garmin, the frequency at which heart rate is measured varies depending on the activity of the user; it also has limited accuracy during swimming (i.e., specific swimming monitors are needed) (Garmin, 2019b, c). The company acknowledges the heart rate assessment can be inaccurate depending on fit of the tracker, type and intensity of the physical activity, and user physical characteristics (Garmin, 2019c). This technique to assess heart rate is also sensitive to large movements, sweat, skin temperature, arrhythmias, health conditions with poor tissue perfusion, and amount of compression when worn (Claes et al., 2017; Gillinov et al., 2017). Garmin suggests several techniques to improve heart rate assessment with photoplethysmography: make sure the watch band is snug against the wrist so it cannot move up and down, wear the watch on the outside of the wrist away from the wrist bone, and avoid wearing it over dark tattoos (Garmin, 2019c).

Sleep

Garmin devices use the accelerometer and heart rate to assess sleep, with newer devices also using additional data such as heart rate variability (Garmin, 2018). For sleep, no studies assessed reliability and only two studies assessed validity for Garmin activity trackers. Both validity studies used a sleep diary for the criterion measure rather than the gold standard polysomnography, which may be why findings were generally poor and variable. Sleep time was overestimated with the Garmin Vivosmart (Lee et al., 2018). When considering other brands of activity trackers, sleep time and sleep efficiency also tended to be overestimated compared to polysomnography due to the lower sensitivity to wake periods, while wake time after sleep onset was underestimated (Baron et al., 2018; Evenson et al., 2015; Kolla et al., 2016).

Among the two Garmin validity studies, one study (Brooke et al., 2017) reported higher validity than the other study (Lee et al., 2018), which may be attributed to their instruction to participants to activate the tracker’s sleep mode function at bedtime. The latter study encouraged the option to confirm sleep times using the Garmin app, but the percent of participants that used this function was not reported (Lee et al., 2018). Nevertheless, the Vivosmart had the lowest MAPE and strongest correlation for total sleep time when compared against seven other trackers under the same conditions (Brooke et al., 2017).

Limitations of Studies

It is important to note that even under the best test conditions, studies of reliability and validity can introduce mis-measurement through several sources of error (Welk et al., 2017). For example, mis-measurement may have introduced some error during data collection of either the gold standard measure or the Garmin tracker. Specification error can occur if the gold standard does not represent the actual concept under study; in this review the threat for this was low in most studies except for sleep assessment.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We identified several specific limitations to the studies we reviewed based on the quality assessment (Appendix 4). First, studies sometimes did not describe whether the Garmin tracker was worn on the dominant or non-dominant wrist. Second, some studies did not describe the settings used or items input on specific trackers, which can make a large difference in findings. Third, studies often did not describe data cleaning, such as whether any outliers were removed. Fourth, it is worth noting that the inclusion criteria varied across studies, with some samples more heterogeneous than others, and at times with limited representativeness. Fifth, many studies did not report on reliability of the trackers, and no study reported both intra- and inter-device reliability. Sixth, we did not identify any studies that reported on physical activity ("active minutes") from the Garmin. Finally, several studies did not account for within-person correlation in their reliability and validity analyses, or did not use appropriate statistical tests for correlated data. Instead, these studies treated each observation as independent, even when multiple observations came from the same person, which can lead to both an under- or over-estimation of agreement (Sainani, 2010).

Limitations of this Review

There were several limitations to this review. We interpreted the CC and MAPE based on prior recommendations uniformly across activity tracker features, even though some features may be more difficult to assess than others. Although the Garmin assesses location and route accuracy (Hallo et al., 2005; Wieters et al., 2012), we did not review the validity of these features since these measures apply to many other devices that the Garmin company offers but were not included in the review of wrist-worn activity trackers. The wearable industry changes quickly and while this review included studies published through 2018, as of June 2019 only one of Garmin activity trackers evaluated (Forerunner 235) was available for purchase from the company's website (Appendix 3). The assumption is that the process used to derive measures, such as energy expenditure and sleep, are stable across types of trackers within the same company. However, this is an unvalidated assumption. Moreover, the precise way the measures are calculated is proprietary and firmware updates can happen without notification, changing the measure attributes over time. For research purposes, it would be ideal for companies to inform users of these changes so that the discontinuity of data collection is avoided.

Conclusions

This systematic review of Garmin activity trackers indicated higher validity of steps, few studies on speed, elevation, and sleep, and lower validity for distance, energy expenditure, and heart rate. This review can facilitate choice in the use of the trackers, as well as to identify gaps in our understanding of its measurement properties. For many features, Garmin offers strategies to improve measurement, such as setting stride length for steps, using a chest strap for heart rate, and using sleep mode for sleep assessment. These strategies were either not mentioned or not tested in many studies. It is anticipated that with the addition of features to Garmin activity trackers used to calculate these metrics, validity should improve.

Similar to reviews of other activity trackers within the same company (Evenson et al., 2015), comparisons between Garmin studies was hampered by the various methodologies and incomplete assessments for a single device type. Specifically, new devices come out before

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

the current ones can be appropriately tested for validity in both laboratory- and field-based settings. Moreover, most activity trackers lack evidence for intra-device and inter-device reliability across most features, indicating the need for further testing and refinement. It is not known when proprietary algorithms change, and what impact those changes have on device features. These challenges, and others, will continue to make it hard to choose an appropriate activity tracker based on its measurement properties until companies become more transparent and researchers more systematically test device features, use the most appropriate comparison measure, and report statistical metrics that appropriately assess the quality of and can be compared across studies. Others propose recommendations for researchers to improve data standardization and harmonization that should be considered to help address deficiencies in the field (Welk et al., 2019).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment:

We gratefully acknowledge funding provided by the North Carolina Translational and Clinical Sciences Institute (National Institutes of Health grant #UL1TR001111). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA (2016). Accuracy of the Vivofit activity tracker. Journal of Medical Engineering and Technology, 40, 298–306. [PubMed: 27266422]

Ammann R, Taube W, Neuhaus M, Wyss T (2016). The influence of the gait-related arm swing on elevation gain measured by sport watches. Journal of Human Kinetics, 51, 53–60. [PubMed: 28149368]

An HS, Jones GC, Kang SK, Welk GJ, Lee JM (2017). How valid are wearable physical activity trackers for measuring steps? European Journal of Sport Science, 17, 360–368. [PubMed: 27912681]

Baron KG, Duffecy J, Berendsen MA, Cheung Mason I, Lattie EG, Manalo NC (2018). Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. Sleep Medicine Reviews, 40, 151–159. [PubMed: 29395985]

Bland J, Altman D (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1, 307–310. [PubMed: 2868172]

Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, Gillan WW, Gusew EE, Kraemer RR (2018). Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc, 50, 624–633. [PubMed: 29189666]

Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM (2017). Concurrent validity of wearable activity trackers under free-living conditions. Journal of strength and Conditioning Research, 31, 1097–1106. [PubMed: 27465631]

Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J (2017). The accuracy of heart rate monitoring by some wrist-worn activity trackers. Annals of Internal Medicine, 166, 610–612.

Cassirame J, Vanhaesebrouck R, Chevrolat S, Mourot L (2017). Accuracy of the Garmin 920 XT HRM to perform HRV analysis. Australasian Physical and Engineering Sciences in Medicine, 40, 831–839. [PubMed: 29058222]

Chen MD, Kuo CC, Pellegrini CA, Hsu MJ (2016). Accuracy of wristband activity monitors during ambulation and activities. Medicine and Science in Sports and Exercise, 48, 1942–1949. [PubMed: 27183123]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, Budts W, Cornelissen V (2017). Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. Journal of Medical Engineering and Technology, 41, 480–485. [PubMed: 28675070]

Crouter SE, Schneider PL, Karabulut M, Bassett DR Jr. (2003). Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. Medicine and Science in Sports and Exercise, 35, 1455–1460. [PubMed: 12900704]

de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M (2016). Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity, 24, 2078–2091. [PubMed: 27670401]

De Vries SI, Van Hirtum HW, Bakker I, Hopman-Rock M, Hirasing RA, Van Mechelen W (2009). Validity and reproducibility of motion sensors in youth: a systematic update. Medicine and Science in Sports and Exercise, 41, 818–827. [PubMed: 19276851]

Dooley EE, Golaszewski NM, Bartholomew JB (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR mHealth and uHealth, 5, e34. [PubMed: 28302596]

Downes MJ, Brennan ML, Williams HC, Dean RS (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open, 6, e011458.

Duking P, Fuss FK, Holmberg HC, Sperlich B (2018). Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR mHealth and uHealth, 6, e102. [PubMed: 29712629]

Duncan MJ, Mummery WK, Dascombe BJ (2007). Utility of global positioning system to measure active transport in urban areas. Medicine and Science in Sports and Exercise, 39, 1851–1857. [PubMed: 17909415]

Ehrler F, Weber C, Lovis C (2016). Influence of pedometer position on pedometer accuracy at various walking speeds: A comparative study. Journal of Medical Internet Research, 18, e268. [PubMed: 27713114]

El-Amrawy F, Nounou MI (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthcare Informatics Research, 21, 315–320. [PubMed: 26618039]

Evenson KR, Goto MM, Furberg RD (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity, 12, 159. [PubMed: 26684758]

Fokkema T, Kooiman TJ, Krijnen WP, Van Der Schans CP, Groot DE (2017). reliability and validity of ten consumer activity trackers depend on walking speed. Medicine and Science in Sports and Exercise, 49, 793–800. [PubMed: 28319983]

Garmin, 2018 New advanced sleep monitoring in Garmin Connect. Accessed on March 27, 2019 at https://www.garmin.com/en-US/blog/fitness/advancedrem/.

Garmin, 2019a Calorie terminology. Accessed on March 27, 2019 at https://support.garmin.com/en-GB/?faq=lkl4cwCLlK7ox362uGQEV7.

Garmin, 2019b Garmin disclaimer: Activity tracking and fitness metric accuracy. Accessed March 27, 2019 at https://www.garmin.com/en-US/legal/atdisclaimer.

Garmin, 2019c Improving the accuracy of the optical heart rate sensor. Accessed March 27, 2019 at https://support.garmin.com/en-US/?faq=xQwjQjzUew4BF1GYcusE59.

Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, Henderzahs-Mason JM, Hagen PT (2018). Determining the validity and accuracy of multiple activity-tracking devices in controlled and free-walking conditions. American Journal of Health Promotion, 32, 1671–1678. [PubMed: 29558811]

Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, Houghtaling P, Javadikasgari H, Desai MY (2017). Variable accuracy of wearable heart rate monitors during aerobic exercise. Medicine and Science in Sports and Exercise, 49, 1697–1703. [PubMed: 28709155]

Gloersen O, Kocbach J, Gilgien M (2018). Tracking performance in endurance racing sports: Evaluation of the accuracy offered by three commercial GNSS receivers aimed at the sports market. Frontiers in Physiology 9, 1425. [PubMed: 30356794]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hallo JC, Manning RE, Valliere W, Budruk M (2005). A case study comparison of visitor self-reported and GPS recorded travel routes. US Dept Agr, Forest Serv Ne Exptl Stn, Radnor. Proceedings of the 2004 Northeatern Recreation Research Symposium, 326,172–177.

Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S (2018). Using fitness trackers and smartwatches to measure physical activity in research: Analysis of Consumer Wrist-Worn Wearables. Journal of Medical Internet Research, 20, e110. [PubMed: 29567635]

Higgins PA, Straub AJ (2006). Understanding the error of our ways: mapping the concepts of validity and reliability. Nursing Outlook, 54, 23–29. [PubMed: 16487776]

Hochsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksass A (2018). Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scandinavian Journal of Medicine and Science in Sports, 28(7), 1818–1827. [PubMed: 29460319]

Hongu N, Orr BJ, Roe DJ, Reed RG, Going SB (2013). Global positioning system watches for estimating energy expenditure. Journal of Strength and Conditioning Research, 27, 3216–3220. [PubMed: 23439338]

Hovsepian D, Meardon SA, Kernozek TW (2014). Consistency and agreement of two devices for running speed. Athletic Training & Sports Health Care, 6, 67–72.

Huang YJ, Xu JK, Yu B, Shull PB (2016). Validity of FitBit, Jawbone UP, Nike Plus and other wearable devices for level and stair walking. Gait and Posture 48, 36–41. [PubMed: 27477705]

International Data Corporation, 2018 Worldwide wearables market ticks up 5.5% due to gains in emerging markets, says IDC. Accessed March 29, 2019 at https://www.idc.com/getdoc.jsp?containerId=prUS44247418, Framingham, MA.

Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA (2016). Validation of biofeedback wearables for photoplethysmographic heart rate tracking. Journal of Sports Science and Medicine, 15, 540–547. [PubMed: 27803634]

Kolla BP, Mansukhani S, Mansukhani MP (2016). Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Review of Medical Devices 13, 497–506. [PubMed: 27043070]

Lamont RM, Daniel HL, Payne CL, Brauer SG (2018). Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait and Posture 63, 104–108. [PubMed: 29729611]

Lee JM, Byun W, Keill A, Dinkel D, Seo Y (2018). Comparison of wearable trackers' ability to estimate sleep. International Journal of Environmental Research and Public Health, 15.

Leth S, Hansen J, Nielsen OW, Dinesen B (2017). Evaluation of commercial self-monitoring devices for clinical purposes: results from the future patient trial, phase I. Sensors, 17.

Madigan EA (2019). Fitness band accuracy in older community dwelling adults. Health Informatics Journal, 25(3), 676–682. [PubMed: 28743215]

Menaspà P, Impellizzeri FM, Haakonssen EC, Martin DT, Abbiss CR (2014). Consistency of commercial devices for measuring elevation gain. International Journal of Sports Physiology & Performance, 9, 884–886. [PubMed: 24338100]

Mooney R, Quinlan LR, Corley G, Godfrey A, Osborough C, Laighin GO (2017). Evaluation of the Finis Swimsense (R) and the Garmin Swim (TM) activity monitors for swimming performance and stroke kinematics analysis. PloS One, 12.

Munck K, Hummeluhr Christensen M, Tahhan A, Dinesen B, Spindler H, Hansen J, Nielsen O, Leth S (2018). Evaluation of self-trackers for use in telerehabilitation. Journal of Usability Studies, 13, 125–137.

Muoio D, 2018 Garmin, ActiGraph partner on wearable-driven medical research, Mobile Health News. Accessed December 12, 2018 at https://www.mobihealthnews.com/content/garmin-actigraph-partner-wearable-driven-medical-research.

Nelson MB, Kaminsky LA, Dickin DC, Montoye AH (2016). Validity of consumer-based physical activity monitors for specific activity types. Medicine and Science in Sports and Exercise, 48, 1619–1628. [PubMed: 27015387]

O'Connell S, G OL, Kelly L, Murphy E, Beirne S, Burke N, Kilgannon O, Quinlan LR (2016). These shoes are made for walking: sensitivity performance evaluation of commercial activity monitors

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

under the expected conditions and circumstances required to achieve the international daily step goal of 10,000 steps. PloS One, 11, e0154956. [PubMed: 27167121]

Plasqui G, Bonomi AG, Westerterp KR (2013). Daily physical activity assessment with accelerometers: new insights and validation studies. Obesity Reviews, 14, 451–462. [PubMed: 23398786]

Pribyslavska V, Caputo JL, Coons JM, Barry VW (2018). Impact of EPOC adjustment on estimation of energy expenditure using activity monitors. Journal of Medical Engineering and Technology, 42, 265–273. [PubMed: 29911930]

Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C (2017). Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. Journal of Medical Engineering and Technology, 41, 208–215. [PubMed: 27919170]

Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, Doyle Iii FJ, Clements MA, Rickels MR, Patton SR, Castle JR, Riddell MC, Jacobs PG (2018). Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR mHealth and uHealth, 6, e10338. [PubMed: 30530451]

Roos L, Taube W, Beeler N, Wyss T (2017). Validity of sports watches when estimating energy expenditure during running. BMC Sports Science, Medicine and Rehabilitation, 9, 22.

Sainani K (2010). The importance of accounting for correlated observations. Physical Medicine and Rehabilitation, 2, 858–861.

Sears T, Avalos E, Lawson S, McAlister IAN, Eschbach C, Bunn J (2017). Wrist-worn physical activity trackers tend to underestimate steps during walking. International Journal of Exercise Science, 10, 764–773.

Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F (2016). Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. Acta Gymnica, 46, 129–135.

Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, Gallagher R (2018). The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. Maturitas, 112, 85–93. [PubMed: 29704922]

Strath SJ, Rowley TW (2018). Wearables for promoting physical activity. Clin Chem, 64, 53–63. [PubMed: 29118062]

Thompson W (2019). Worldwide survey of fitness trends for 2019. ACSM Health Fitness J, 22, 10–18.

Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C (2017). Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Physical Therapy, 97, 581–588. [PubMed: 28339904]

Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A (2006). Evaluation of quality of commercial pedometers. Canadian Journal of Public Health, 97 Suppl 1, S10–15.

Wahl Y, Duking P, Droszez A, Wahl P, Mester J (2017). Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Frontiers in Physiology, 8, 725. [PubMed: 29018355]

Wallen MP, Gomersall SR, Keating SE, Wisloff U, Coombes JS (2016). Accuracy of heart rate watches: Implications for weight management. PLoS One, 11, e0154420. [PubMed: 27232714]

Wang L, Liu T, Wang YH, Li QQ, Yi JG, Inoue Y (2017). Evaluation on step counting performance of wristband activity monitors in daily living environment. IEEE Access, 5, 13020–13027.

Welk G, Morrow J, Saint-Maurice P (2017). Measures Registry User Guide: Individual Physical Activity. Accessed at http://nccor.org/tools-mruserguides/wp-content/uploads/2017/NCCOR_MR_User_Guide_Individual_PA-FINAL.pdf. National Collaborative on Childhood Obesity Research, Washington D.C.

Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L (2019). Standardizing analytic methods and reporting in activity monitor validation studies. Medicine and Science in Sports and Exercise, 51, 1767–1780. [PubMed: 30913159]

Wieters KM, Kim JH, Lee C (2012). Assessment of wearable global positioning system units for physical activity research. Journal of Physical Activity and Health, 9, 913–923. [PubMed: 21975729]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Woodman JA, Crouter SE, Bassett DR Jr., Fitzhugh EC, Boyer WR (2017). Accuracy of Consumer Monitors for Estimating Energy Expenditure and Activity Type. Medicine and Science in Sports and Exercise, 49, 371–377. [PubMed: 27580155]

Wright SP, Hall Brown TS, Collier SR, Sandberg K (2017). How consumer physical activity monitors could transform human physiology research. American Journal of Physiology, 312, R358–R367. [PubMed: 28052867]

Yavelberg L, Zaharieva D, Cinar A, Riddell MC, Jamnik V (2018). A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: A novel approach. Journal of Diabetes Science and Technology, 12(3), 569–576. [PubMed: 29320885]

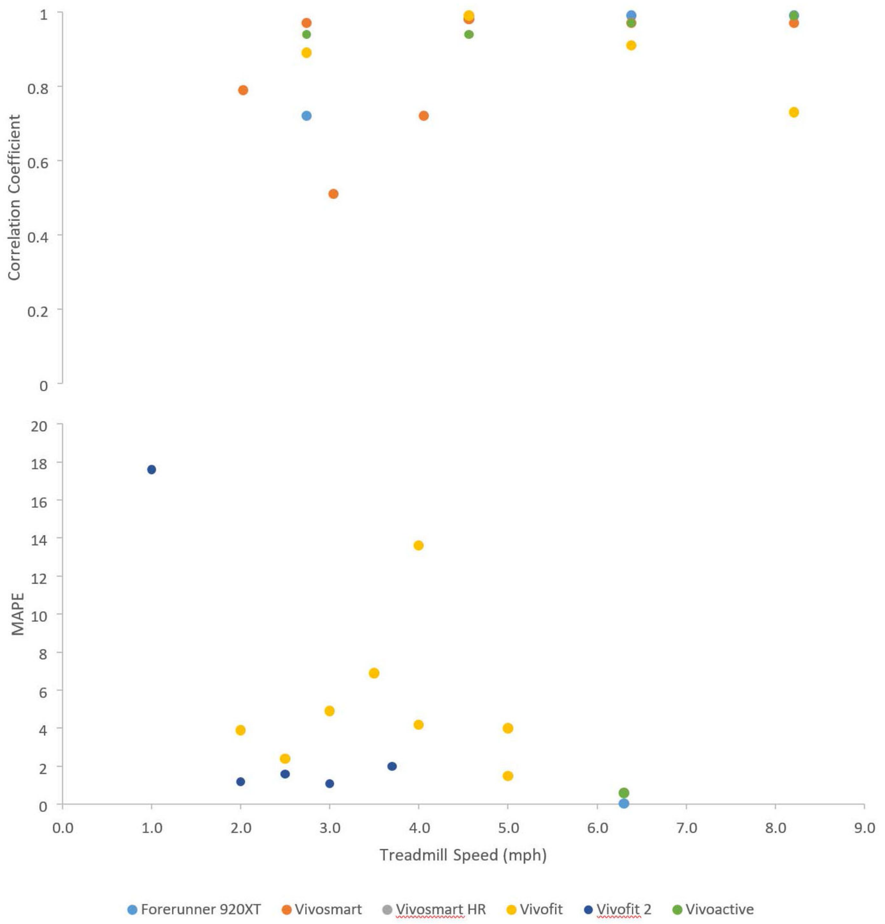

Figure 1:

Correlation coefficients and mean absolute percentage error (MAPE) for steps taken on the treadmill at zero percent grade measured with Garmin activity trackers

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

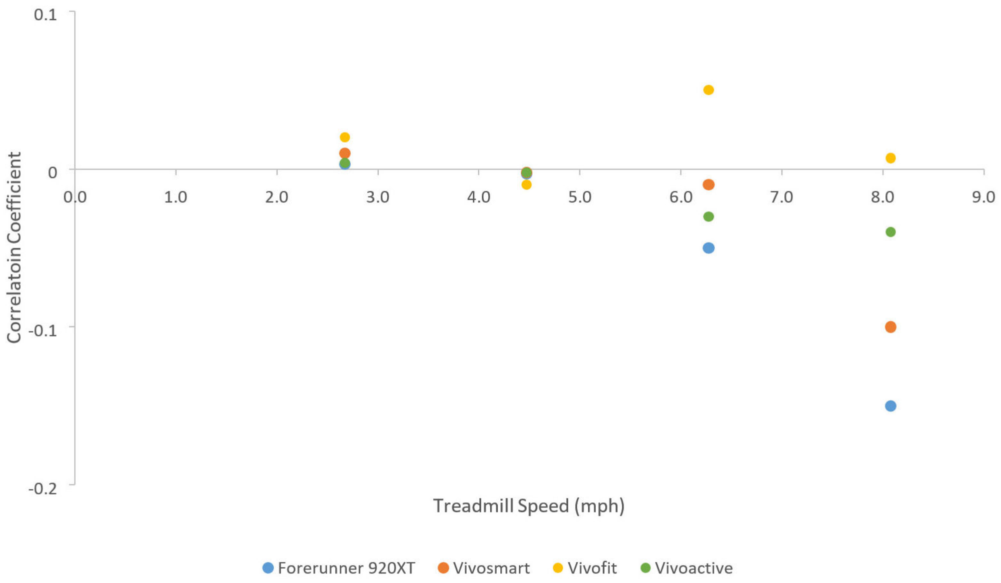

Figure 2:

Correlation coefficients for distance taken on the treadmill at zero percent grade measured with Garmin activity trackers

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

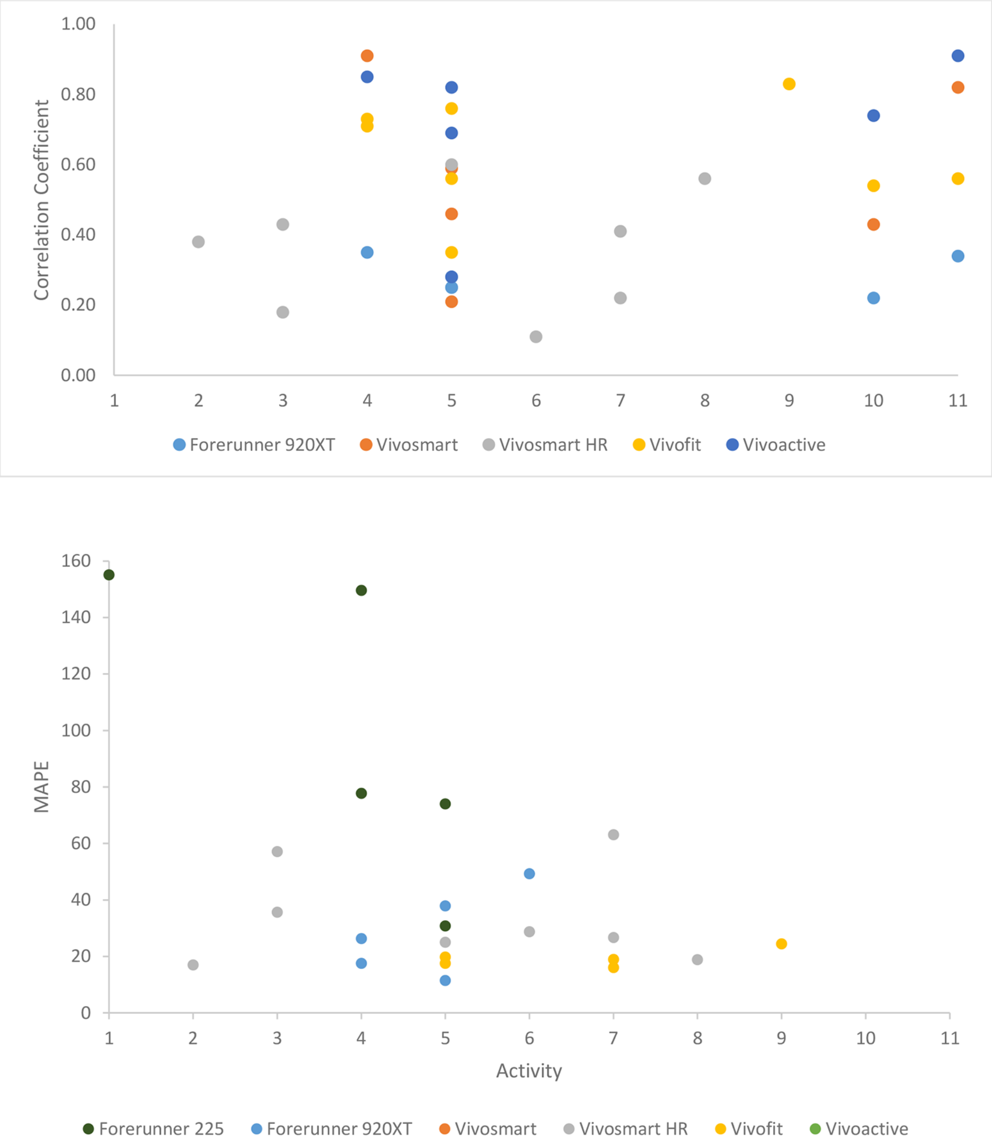

Figure 3:

Correlation coefficients and mean absolute percentage error (MAPE) for energy expenditure measured with Garmin activity trackers

Footnote: Rest= 1; Activities of daily living= 2; Resistance training= 3; Walking=4; Running= 5; Running maximal= 6; Cycling=7; Cycling maximal= 8; Two days of wear= 9; Intermittent activity= 10; Outdoor activity= 11

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

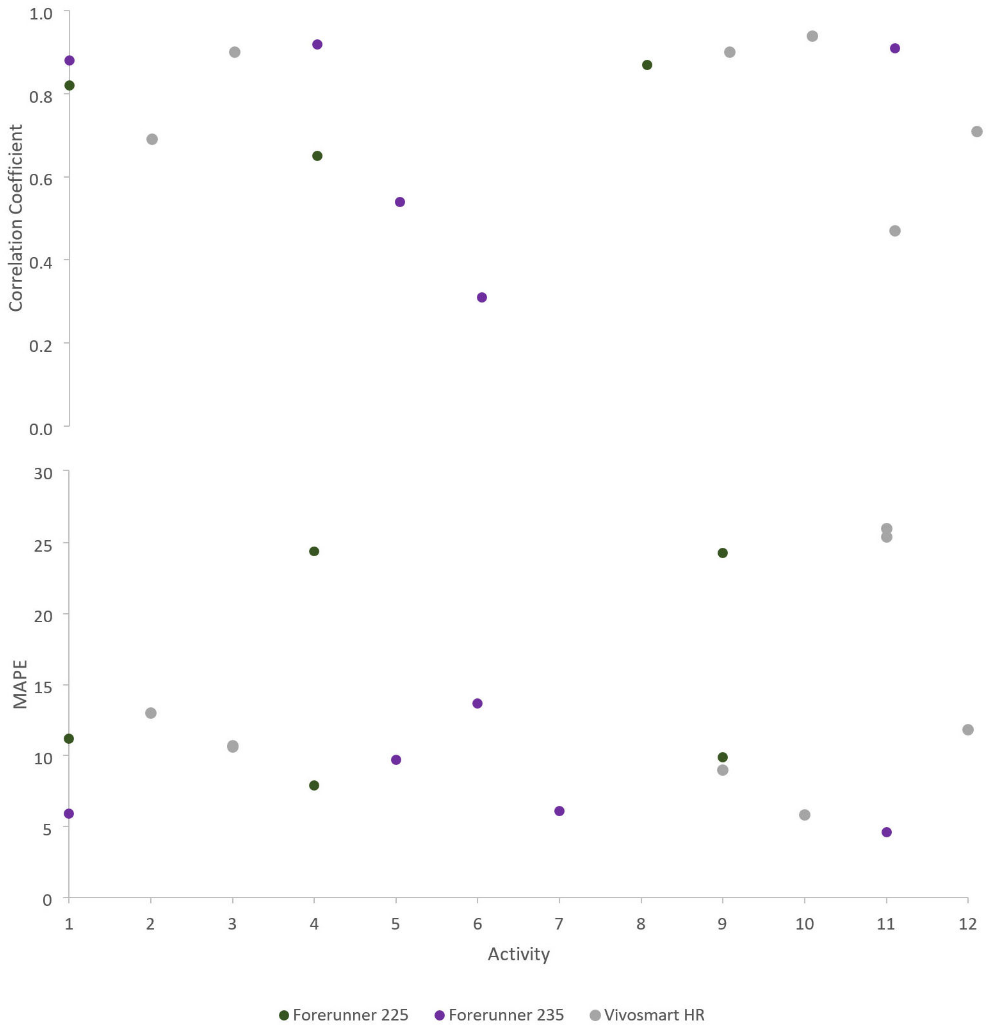

Figure 4:

Correlation coefficients and mean absolute percentage error (MAPE) for heart rate measured with Garmin activity trackers

Footnote: Rest= 1; Activities of daily living= 2; Resistance training= 3; Walking=4; Elliptical (no arms)= 5; Elliptical (with arms)= 6; Treadmill= 7; Walking with grade= 8; Running= 9; Running maximal= 10; Cycling= 11; Cycling maximal= 12

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Garmin studies of reliability and validity (listed by author's last name and publication year)

Garmin Device	Validity or Reliability	Steps	Distance	Speed	Elevation	Energy Expenditure	Heart Rate	Sleep
Forerunner 225	Validity					Dooley 2017	Claes 2017; Dooley 2017	
Forerunner 235	Validity						Gillinov 2017	
Forerunner 305	Validity			Hovsepian 2014		Hongu 2013		
	Reliability			Hovsepian 2014				
Forerunner 310XT	Validity				Menaspa 2014			
Forerunner 910XT	Validity				Ammann 2016			
Forerunner 920XT	Validity	Wahl 2017	Wahl 2017			Roos 2017; Wahl 2017		
Vivoactive	Validity	Wahl 2017	Wahl 2017			Wahl 2017		
Vivofit	Validity	Alsubheen 2016; An 2017; Chen 2016; Ehrler 2016; El-Amrawy 2015; Huang 2016; O'Connell 2016; Simunek 2016; Wahl 2017	Huang 2016; Wahl 2017			Alsubheen 2016; Brooke 2017; Pribyslavska 2018; Price 2017; Wahl 2017; Woodman 2017		Brooke 2017
	Reliability	Chen 2016; O'Connell 2016						
Vivofit 2	Validity	Gaz 2018; Hochsmann 2018; Leth 2017; Munck 2018; Wang 2017	Gaz 2018			Yavelberg 2018		
Vivosmart, Vivosmart HR, and Vivosmart HR +	Validity	Fokkema 2017; Sears 2017; Wahl 2017	Wahl 2017			Boudreaux 2018; Reddy 2018; Wahl 2017	Boudreaux 2018; Reddy 2018	Lee 2018
	Reliability	Fokkema 2017						

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Characteristics of studies included in the systematic review (listed by author's last name and publication year)

Author (Year)	Location of Lab or Recruitment Area	Sample Size for Validity and Reliability Studies*	% Female	Mean Age (SD), Range	Mean body mass index (SD), range in kilograms/ meters squared	Data Collection Period	Inclusion Criteria	Features Tested	Number Garmin Features Tested	Number Devices Tested**
Alsubheen (2016)	Newfoundland, Canada	13 (V)	38	40 (11.9)	27.0 (3.4)	2015	Apparently healthy adult using PARQ as a screener	EE, S	2	1
Ammann (2016)	Switzerland	3 (V)	0	25.5 (1.3)	not reported	not reported	Recreational runners, practicing endurance sports more than 300 minutes/week with differing heights	E	1	3
An (2017)	Omaha, Nebraska, USA	35 (V)	51	31.0 (11.8), 19–65	23.8 (3.1)	not reported	Apparently healthy, completed PARQ, able to walk/run safely on treadmill and around an indoor track, does not use a walking aid, not pregnant, does not have an implanted electromagnetic device	S	1	10
Boudreaux (2018)	Hammond, Louisiana, USA	50 (V)	56	Females 22.7 (3.0), Males 22.0 (2.7), All 18–35	Females 25.8 (4.8), Males 27.1 (3.6)	October 2015 to June 2016	No cardiovascular disease or musculoskeletal injury within the past 6 months	EE, HR	2	8
Brooke (2017)	Omaha, Nebraska, USA	95 (V)	64	28.5 (9.9), 19–60	25.7 (3.4), 17–34.3	not reported	Able to perform activites of daily living without limitations, completed PARQ, does not require walking aids or have walking impairments	EE, SL	2	8
Chen (2016)	Kaohsiung City, Taiwan	30 (V and R)	50	21.5 (2.0)	21.5 (1.9)	February 2015 to May 2015	At least 20 years old, normal body mass index, could ambulate without assistance, normal gait pattern	S	1	3
Claes (2017)	Leuven, Belgium	12 (V)	50	28.0 (4.79)	22.14 (3.46)	October 2015 to April 2016	Regularly physically active men or women between 20–40 years of age and no known musculoskeletal pathology or cardiovascular, respiratory or metabolic disease.	HR	1	1
Dooley (2017)	Austin, Texas, USA	62 (V)	58	22.6 (4.3), 18–38	24.6 (4.8), 17.1–45.0	not reported	Caffeine free for 12 hours, fasted for 3 hours, non-smoker, no disability contraindicated for exercise, and no tattoos, piercings, or braces where device would be worn	EE, HR	2	3

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author (Year)	Location of Lab or Recruitment Area	Sample Size for Validity and Reliability Studies*	% Female	Mean Age (SD), Range	Mean body mass index (SD), range in kilograms/ meters squared	Data Collection Period	Inclusion Criteria	Features Tested	Number Garmin Features Tested	Number Devices Tested**
Ehrler (2016)	Geneva, Switzerland	21 (V)	57	34.5 (15.7)	not reported	not reported	Healthy volunteers, able to walk at least 500m and not have any walking disability	S	1	4
El-Amrawy (2015)	Alexandria, Egypt	4 (V)	0	26.5 (12.8)	not reported	March 2014-June 2015	Apparently healthy adult 22–36 years	S	1	17
Fokkema (2017)	Groningen, The Netherlands	30 to 31 (V and R)	48	32 (12)	22.6 (2.4)	Fall 2016	Apparently healthy adult volunteers	S	1	10
Gaz (2018)	Rochester, Minnesota, USA	32 (V)	69	36 (8), 26–56	26.8 (5.2), 18.2–41.8	not reported	No known orthopedic limitations, no absolute contridictions to physical activity, employees of the institution	D, S	2	6
Gillinov (2017)	Cleveland, Ohio, USA	25 (V)	54	38 (12)	25 (3.5)	June 2016 to August 2016	At least 18 years old, could safely perform an 18 minute exercise protocol, and no known cardiovascular or lung disease, presence of cardiac pacemaker, treatment with beta-blockers or heart rhythm medications, and self-reported chest pain, dizziness, or loss of balance	HR	1	6
Hochsmann (2018)	Basel, Switzerland	20 (V)	Group 1: 60; Group 2: 80	Group 1: 22, 21–23; Group 2: 53, 52–66	Group 1: 23, 21–25; Group 2: 24, 22–29	January 2017 to March 2017	Apparently healthy volunteers	S	1	7
Hongu (2013)	Tucson, Arizona, USA	16 (V)	56	Females: 22.6 (2.6); Males: 21.3 (1.5)	Females: 22.5 (1.4); Males: 22.8 (2.3)	not reported	Apparently healthy college students free from cardiovascular or metabolic diseases or physical impariments that would interfere with walking	EE	1	4+1
Hovsepian (2014)	La Crosse, Wisconsin, USA	13 (V and R)	not reported	25.3 (2.5)	not reported	not reported	At least 18 years and self-reported running an average of >=10 miles/ week during the past year	S	1	2
Huang (2016)	Shanghai, China	40 (V)	25	23.9 (2.8)	21.4 (2.5)	September 2014 to October 2014	>18 years old, able to walk on flat ground for at least 10 minutes and up or down stairs for at least 6 minutes continuously, body mass index <32 kg/m2, and no previous history of injury or disease inhibiting normal gait	D, S	2	5
Lee (2018)	Omaha, Nebraska, USA	40 (V)	54	27.6 (11.0), 19–66	25.3 (4.6), 19.4–39.7	not reported	>=19 years old, no insomnia	SL	1	6

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author (Year)	Location of Lab or Recruitment Area	Sample Size for Validity and Reliability Studies*	% Female	Mean Age (SD), Range	Mean body mass index (SD), range in kilograms/ meters squared	Data Collection Period	Inclusion Criteria	Features Tested	Number Garmin Features Tested	Number Devices Tested**
Leth (2017)	Aalborg, Denmark	22 (V)	50	31.1 (8.0), 22–52	not reported	November 2015 to June 2016	No walking disabilities that could lead to unnatural walking patterns	S	1	5+1
Menaspa (2014)	Varese, Italy (study 2)	1 (V)#	not reported	not reported	not reported	September 2012	not reported	E	1	4
Munck (2018)	Aalborg, Denmark	22 (V)	50	27 (7.25), 21–49	25.0 (3.8), 20.1–36.4	November 2015	>18 years old, capable of understanding Danish, did not suffer from previous neurologic, musculoskeletal, or mental illness, no use of walking aids, not pregnant	S	1	6
O’Connell (2016)	Galway, Ireland	15 (V)	53	21.1 (1.1)	Females 21.9 (1.8), Males 23.6 (2.7)	February 2015 to July 2015	No history of cardiovascular disease or neurological disorder	S	1	4
Pribyslavska (2018)	Murfreesboro, Tennessee, USA	34 (V)	32	25.8 (4.9)	24.4 (4.4)	Fall 2016-Spring 2017	classified as either low or moderate risk according to the American College of Sports Medicine cardiovascular risk classification, physically active	EE	1	3
Price (2017)	Melbourne, Australia	14 (V)	21	23.0 (6.0)	22.8 (2.6)	September 2014 to September 2015	Able to walk and run continuously on a treadmill unaided, healthy and free of factors associated with exercise risk as determined through standard screening procedures	EE	1	3
Reddy (2018)	Portland, Oregon, USA and Toronto, Canada	20 (V)	55	27.5 (6.0)	22.5 (2.3)	December 2017-February 2018	Healthy adults, screening used PARQ	EE, HR	2	2+1
Roos (2017)	Magglingen, Switzerland	20 (V)	40	23.9 (1.9)	not reported	January 2016 to March 2016	Recreational or competitive runner, no injury to lower extremities within past year.	EE	1	3
Sears (2017)	Buies Creek, North Carolina, USA	10 (V)	50	23.3 (5.2), 18–40	not reported	Spring 2016	Recreationally active, low- or moderate-risk for cardiovascular disease	S	1	5
Simunek (2016)	Olomouc, Czech Republic	20 (V)	30	34.0 (6.3), 25–52	24.3 (4.0)	December 2014 to February 2015	No history of injury or illness affecting mobility	S	1	2+2
Wahl (2017)	Cologne, Germany	20 (V)	50	Females: 24.2 (1.9), Males: 26.1 (2.8)	not reported	not reported	Apparently healthy and active sport students	D, EE, S	3	11

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author (Year)	Location of Lab or Recruitment Area	Sample Size for Validity and Reliability Studies*	% Female	Mean Age (SD), Range	Mean body mass index (SD), range in kilograms/ meters squared	Data Collection Period	Inclusion Criteria	Features Tested	Number Garmin Features Tested	Number Devices Tested**
Wang (2017)	Hangzhou, China	9 (V)	44	22.0 (1.0)	not reported	Spring 2015	Apparently healthy participants	S	1	7
Woodman (2017)	Knoxville, Tennessee, USA	28 (V)	29	25.5 (3.7)	24.9 (2.6)	January 2015 to May 2015	Completed PARQ, not currently pregnant, obese, or have orthopedic or musculoskeletal issues that would limit activity, able to run on treadmill for 5 min at 134.1m/min with 0% grade	EE	1	5
Yavelberg (2018)	Toronto, Ontario, Canada	25 (V) but smaller sample wore Garmin	44	25.0 (7.6), 18–55	Females without diabetes 23.8 (2.7), Females with diabetes 24.7 (1.5), Males without diabetes 26.0 (2.2), Males with diabetes 23.6 (2.2)	2014 to 2016	At least 16 years, otherwise healthy and active with moderate to high levels of physical activity. Eight participants had a diagnosis of type 1 diabetes.	EE	1	5

Abbreviations: D, distance; E, elevation; EE, energy expenditure; HR, heart rate; PARQ, Physical Activity Readiness Questionnaire; R, reliability;SL, sleep; SD, standard deviation; S, steps; V, validity

*The sample size was based from the article on the number who tested a Garmin device. For some studies, this was less than the full sample described for gender, age, and body mass index.

**The number of devices tested is listed as two numbers if the gold standard assessment included a device (e.g., accelerometer, activity tracker).

#Only results from study 1 were included, since study 2 did not use an activity tracker.