



## Article

# Validity of Wearable Monitors and Smartphone Applications for Measuring Steps in Semi-Structured and Free-Living Settings

Manolis Adamakis

School of Physical Education and Sport Science, National and Kapodistrian University of Athens, 17237 Dafne, Greece; manosadam@phed.uoa.gr

**Abstract:** Wearable technologies have become powerful tools for health and fitness and are indispensable everyday tools for many individuals; however, significant limitations exist related to the validity of the metrics these monitors purport to measure. Thus, the purpose of the present study was to validate the step count of three wearable monitors (i.e., Yamax 3D Power-Walker, Garmin Vivofit 3 and Medisana Vifit), as well as two Android apps (i.e., Accupedo Pedometer and Pedometer 2.0), in a sample of healthy adults. These monitors and apps were evaluated in a lab-based semi-structured study and a 3-day field study under habitual free-living conditions. A convenience sample of 24 healthy adults (14 males and 10 females;  $32.6 \pm 2.5$  years) participated in both studies. Direct step observation and Actigraph served as the criterion methods and validity was evaluated by comparing each monitor and app with the criterion measure using mean absolute percentage errors (MAPE), Bland–Altman plots, and Intraclass Correlation Coefficients. The results revealed high validity for the three wearable monitors during the semi-structured study, with MAPE values approximately 5% for Yamax and Vifit and well below 5% for Vivofit, while the two apps showed high MAPE values over 20%. In the free-living study all monitors and apps had high MAPE, over 10%. The lowest error was observed for Yamax, Vifit and Pedometer app, while Accupedo app had the highest error, overestimating steps by 32%. The present findings cannot support the value of wearable monitors and apps as acceptable measures of PA and step count in free-living contexts. Wearable monitors and apps that might be valid in one context, might not be valid in different contexts and vice versa, and researchers should be aware of this limitation.



**Citation:** Adamakis, M. Validity of Wearable Monitors and Smartphone Applications for Measuring Steps in Semi-Structured and Free-Living Settings. *Technologies* **2023**, *11*, 29. <https://doi.org/10.3390/technologies11010029>

Academic Editor: James Navalta

Received: 10 January 2023

Revised: 3 February 2023

Accepted: 10 February 2023

Published: 13 February 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** validity; accuracy; accelerometer; physical activity measurement; free-living monitoring; wearable monitor; activity tracker; consumer-level monitor; smartphone app; step count

## 1. Introduction

The health benefits of a physically active lifestyle during the lifespan are well documented. These include improved cardiorespiratory and muscular fitness, bone and cardiometabolic health, a lower risk for a series of major diseases (e.g., hypertension and diabetes), improved mental health and positive effects on weight status [1]. To monitor and set physical activity (PA) goals step counting is widely implemented [1]. Steps are a basic unit of locomotion and, as such, provide an easy-to-understand metric of ambulation, which is an important component of daily PA. Recently large cohort studies have used step count to estimate how PA is associated with mortality risk. Saint-Maurice et al. [2] concluded that a greater number of daily steps was significantly associated with lower all-cause mortality. A similar conclusion was supported by Paluch et al. [3], who further showed that there was a progressively decreasing risk of mortality among adults aged 60 years and older with increasing number of steps per day until 6000–8000 steps per day. The CARDIA study found that among men and women in middle adulthood, participants who took approximately 7000 steps/day or more experienced lower mortality rates compared with participants taking fewer than 7000 steps/day [4], and for each 1000 daily step

count increase at baseline, risk reductions in all-cause mortality (6–36%) and cardiovascular disease (5–21%) at follow-up were estimated [5].

Currently, the self-assessment of steps can be accomplished through wearable, easily obtainable technology such as pedometers, smartphones, and PA trackers. Unlike the measure of moderate-to-vigorous PA minutes per week, the step counts metric provides a comparable measure to how caloric intake in most dietary guidance is standardized [1].

Recent survey data of fitness trends between 2020 and 2022 showed that wearable technology was the number one most popular trend in 2020 and 2022 [6,7], and the second most popular trend in 2021, behind online training [8]. In 2021, global shipments of wearables, watches, wristbands, and other wearables stood at 533.6 million units. There was a 20% year-over-year growth indicating a growing market [9]. Regarding smartphones, in 2021, around 1.43 billion smartphones were sold worldwide. Less than half of the world's total population owned a smart device in 2016; however, the smartphone penetration rate has continued climbing, reaching 78% in 2020 [10].

As a result, the use of wearable monitors and smartphone applications (apps) to estimate steps per day would provide a useful tool for researchers and the public to address a variety of health and PA issues. Measuring step counts has been shown to motivate diverse samples of individuals to increase their daily PA [1]. Interventions using apps or wearable activity monitors seem to be effective in promoting PA and may lead to an average increase of 1850 steps per day, an amount that is known to have significant clinical impact in reducing mortality risk. The apps and trackers seem to work best when complemented by personalisation or text-messaging [11]. In addition, the Physical Activity Guidelines Advisory Committee [1] found that there is strong evidence that wearable activity monitors, including step counters (pedometers) and accelerometers, when used in conjunction with goal-setting and other behavioral strategies, can help increase PA in the general population of adults as well as in those who have type 2 diabetes. On the other hand, moderate evidence indicates that mobile phone programs consisting of, or including, text-messaging have a small to moderate positive effect on PA levels in general adult populations [1].

While the increasing acceptance and use of these monitors and apps has resulted in a surge of validation studies, accurate assessment of PA remains challenging. Much of the published research fails to rigorously evaluate validity, and there is a lack of consistency across the published protocols, limiting valid comparisons between monitors and apps. Such validation studies are typically performed in a laboratory or field-based context, not both. Measurement accuracy is indispensable when tracking PA variables to provide meaningful measures of PA [12].

A literature review of reviews on techniques for PA measurement in adults found that, for step counting, activity monitors and pedometers achieved high levels of criterion validity. When comparing the two, pedometers appeared to be less accurate than monitors, tending to underestimate steps when compared to direct observation [13]. Another systemic review examining 158 publications and 45 monitors concluded that wearable monitors are accurate for measuring step count in the laboratory but exhibit a wider range of inaccuracy in free-living environments [14]. Regarding the validity of mobile apps to count steps, a literature review showed conflicting evidence. Apps tended to be less accurate at lower velocities and when the smartphone was carried near the hip (e.g., pocket trousers). Additionally, studies conducted in free-living environments found significant errors higher than 10%, suggesting that the apps tested were not valid for counting steps in day-to-day activities [15].

In free-living conditions, recent studies found conflicting evidence of step count validity. Ferguson et al. [16] concluded that the consumer-level activity monitors in their study showed strong validity for the measurement of steps; however, validity for each construct ranged widely, with the Fitbit One, Fitbit Zip and Withings Pulse being the strongest performers. Breteler et al. [17] came to a similar conclusion, since validity varied widely between monitors, with the Apple Watch being the most accurate and Yamax Digiwalker the least accurate for step count in free-living conditions. In another field study

of healthy individuals, all but one of the activity monitors showed a substantial correlation with the criterion device and Mean Absolute Percentage Error (MAPE) lower than 10%. However, at slower speeds in the lab-based study, the accuracy of all monitors substantially deteriorated [18].

On the other hand, the MAPE for the total step count during a 3-day study was high with a general underestimation of steps by all monitors of more than 20% compared to the criterion measure [19]. Bai et al. [20] also found high MAPE in a study comparing three activity monitors (i.e., Fitbit Charge 2, Fitbit Alta, and Apple Watch 2) during a 24-h free-living condition, and these MAPE ranged from 17.1% to 35.5%. Similarly, high MAPE values were estimated in a study comparing iPhone step count with a validated pedometer. The largest underestimation of steps by the iPhone was observed among those who reported to have seldom carried their iPhones [21].

Wearable technologies have become powerful tools for health and fitness and indispensable everyday tools for many individuals; however, significant limitations exist related to the validity of the metrics these monitors purport to measure [22]. Since there is an apparent potential for these monitors and apps to measure and promote PA [23] and, due to the conflicting validity evidence that currently exists, there is a need to carry out more studies of high methodological quality. Thus, the purpose of the present study was to validate the step count of three wearable monitors, as well as two Android apps, in a sample of healthy adults. Based on the evaluation framework proposed by Keadle et al. [24], 2 validation studies were implemented: a semi-structured (lab-based) and a naturalistic (free-living) one.

## 2. Materials and Methods

Three wearable activity monitors and two smartphone apps were evaluated in a lab-based semi-structured study and a 3-day field study under habitual free-living conditions. The studies were reviewed and approved by the Social Research Ethics Committee of University College Cork in Ireland, and they were conducted according to the principles of the Declaration of Helsinki. Participants were informed about all relevant aspects (e.g., risks and benefits) of the studies before enrolling, were notified about the right to refuse to participate or to withdraw consent at any time without reprisal, and then provided written informed consent.

### 2.1. Participants

Due to insufficient data regarding the validity of the monitors and apps to detect steps in a free-living setting, the sample size based on the expected ICC was calculated [25]. A systematic review of comparable validation studies has shown that the ICC is usually well above 0.7 [26], and this is confirmed by two recent studies [18,19]. With a significance level of 0.05, the sample size needed to attain a conservatively calculated ICC of 0.6 at a targeted power of 80% was determined as 17 participants. Accounting for a possible dropout of 10%, the aim was to include at least 20 participants in the study. Thus, a convenience sample of 24 healthy, normal weight adults ( $n = 14$  males,  $n = 10$  females; age range 25–35 years) with typical gait, no contraindications for exercise and no known orthopedic limitations or other physical limitations that would prevent them from completing the assessments, were recruited, and participated in both studies (with no dropouts).

### 2.2. Anthropometric Assessment

Standing height was measured to the nearest 0.1 cm using a wall mounted Harpenden stadiometer (Harpenden, London, UK) using standard procedures. Body mass was measured with participants in light clothes and bare feet on an electronic scale (Omron BF-511) to the nearest 0.1 kg. Body mass index (BMI) was calculated as weight (kg) / height squared ( $m^2$ ). Regarding step length estimation, the INTERLIVE network definition of a step was used [27]: “The act of raising one foot and putting it down in another spot, resulting in the displacement of the centre of mass” (p. 788). The average walking step

length was calculated by performing 20 normal steps and measuring the distance between the start and end line, then dividing the total distance by 20 steps. The same procedure was followed to calculate running step length. All anthropometric measurement results are presented in Table 1.

**Table 1.** Participants' characteristics (Mean  $\pm$  SD).

	Males ( <i>n</i> = 14)	Females ( <i>n</i> = 10)	Total ( <i>n</i> = 24)
	M (SD)	M (SD)	M (SD)
Age (years)	32.6 (2.5)	29.6 (3.5)	31.3 (3.3)
Weight (kg)	81.8 (4.0)	63.5 (5.9)	74.1 (10.4)
Height (m)	1.87 (0.04)	1.66 (0.06)	1.78 (0.11)
BMI (kg/m <sup>2</sup> )	23.4 (1.1)	23.1 (1.5)	23.3 (1.2)
Walking step length (cm)	82.4 (5.2)	66.8 (5.0)	75.8 (9.3)
Running step length (cm)	104.2 (2.7)	86.6 (4.4)	96.8 (9.5)

### 2.3. Wearable Activity Monitors

Three wearable activity monitors were evaluated: Yamax EX510 3D Power-Walker (Yamax; Yamasa Tokei Keiki Co., Ltd., Tokyo, Japan), Garmin Vivofit 3 (Vivofit; Garmin Ltd., Schaffhausen, Switzerland) and Medisana Vifit (Vifit; Medisana AG, Neuss, Germany).

Yamax is a low-cost waist-worn accelerometer that works using a piezoelectric sensor. With inbuilt 3D axis technology, it can accurately measure data at almost any angle while being around the waist, in the pocket or handbag. It counts steps taken, distance travelled, and calculates calories burned. It does not have specific software for assessing data; however, data can be stored in the in-built memory for 30 days.

Vivofit is a mid-cost wrist-based, triaxial accelerometer-based monitor that measures steps taken, distance travelled, calories expended and sleep quality. When paired with a Garmin heart rate chest strap, the monitor can also measure the user's heart rate and incorporate this measurement in the EE estimation algorithm. The Garmin Connect software was used to assess step data for Vivofit.

Vifit is a low-cost waist-worn accelerometer that counts and keeps track of steps taken and calories burned. By means of a triaxial accelerometer and altimeter technology, Vifit records PA. In comparison to more sophisticated activity monitors, it only has the option to insert walking step length (instead of both walking and running). Vifit also measures the duration and quality of sleep. The VitaDock Online software was used to assess step data.

### 2.4. Accelerometer-Based Apps

This study used one Samsung Galaxy S8, based on the Android 10.1 operating system. Inclusion criteria for all apps were retrieved from previous protocols [28,29]: (1) free of charge indefinitely after download, applications with a free trial period of finite length were excluded; (2) full and efficient functionality after downloading, without additional software download being necessary; (3) functionality only through the built-in accelerometer (no GPS or 4G/5G signal); (4) ability to record the number of steps taken, average speed, total distance, and energy expenditure; (5) manual input of demographic and anthropometric data (sex, age, weight, height, and step length for walking and running); (6) manual choice of activity type (i.e., walking or running); (7) among the most popular and downloadable applications, according to users' ratings and number of downloads from the Google Play Store.

Based on the previously described criteria, two accelerometer-based apps were selected: Accupedo Pedometer (Accupedo; Corusen LLC, Keller, TX, USA) and Pedometer 2.0 (Pedometer; DSN Inc., Tokyo, Japan).

Accupedo is a pedometer app that monitors daily walking and calculates the PA level. The accuracy of this app is based on triaxial motion recognition algorithms which tracks

walking patterns by filtering and rejecting non-walking activities. In addition, this app has enough display modes such as steps, distance, minutes, and calories.

Pedometer counts steps, calories, distance, speed, average speed, time in motion, takes all sorts of graphics and splits table in different modes, according to BMI. Furthermore, it has the unique feature of self-calibration capability, which was used to determine the appropriate sensitivity settings for every participant separately.

### *2.5. Lab-Based Semi-Structured Study*

To evaluate the validity of activity monitors and smartphone apps during normal walking speeds, a lab-based semi-structured study of 400 steps was conducted. The participants were fitted with three activity monitors and one smartphone, which was running simultaneously the two apps. Vivofit was worn on the left wrist. Yamax and Vifit as well as the smartphone, were strapped close to the body on a waist-worn elastic belt over the left hip, near the anterior axillary line, and were counterbalanced for anterior and posterior placement on the hip among participants. All devices were updated with the participants' age, sex, height, dominant hand, weight, and step length. All monitors' firmware and apps' software were updated to the latest available version.

In the lab-based testing participants had to walk a total of 400 steps at a self-selected pace. During walking, participants ascended and descended 20 stairs (height = 15.8 cm, depth = 32.0 cm) located inside a building stairwell. The stair height and depth were selected to be similar to previous stair walking validation studies [30]. Participants first ascended the stairs, then rested for 30 s, then descended and rested for another 30 s, and finally completed the remaining test steps.

The criterion measure for steps was two manual counters who objectively measured steps with the use of a hand-held counter device (GOGO Four Digit Hand Tally Counter, atafa.com). The researchers observed the leg movement of the participants and were separated so they could not view each other's thumb motion nor hear the "clicking" from the counter device. This prevented any synchronized counting between the two. On the rare occasions when their observations were not in agreement (the difference was never greater than one step), the greater of the two values was recorded.

### *2.6. Free-Living Field Study*

To explore the validity of activity monitors under free-living conditions, a 3-day field study was conducted. The timeframe of three days was selected as it seemed reasonable to expect that all typical activities of daily living would be performed by the participants during that timeframe if they were, in fact, typical activities [18,19]. A longer timeframe would certainly have provided even more data but would also likely have affected participants' compliance in accurately recording all activities in the diary. In this study, both criterion validity, with step count recorded by Actigraph wGT3X-BT (ActiGraph, Pensacola, FL, USA) as the criterion, and concurrent validity were examined.

The monitors and smartphone fitting procedure of the semi-structured study was implemented. In addition, the Actigraph wGT3X-BT was fitted to the participants and was worn at the waist on the right side, using the elastic belt provided by the manufacturer, and was positioned in line with the armpit and knee with the USB port cover facing up. The device was operated according to the manufacturer's default settings (i.e., sampling rate of 30 Hz). ActiLife 6 (v6.13.3) software (ActiGraph, Pensacola, FL, USA) was later used to reintegrate data to 60-s epochs and calculate daily step count. Actigraph was used as the criterion as it is a reliable and valid tool that has been widely used in various populations and is one of the most frequently used criterion measurement to validate other monitors in research setting [19,20,31,32].

The participants were initially asked to remove and re-attach the devices to familiarize themselves with the routine under the supervision of the researchers and to prove that they were capable of adhering to the protocol. They were then instructed to place the devices on their body directly after getting up in the morning and to wear them simultaneously

during waking hours, except during bathing and water-based activities, and return them 3 days later. If the devices had to be taken off throughout the day, participants were further instructed to always put on and take off all devices at the same time. In addition, they were asked to record the wear time of all devices as well as the times awake for each day in a diary and to adhere to normal daily activities. Upon returning the devices, the diary records were discussed and asked the participants specifically about periods when the devices were not worn simultaneously. Subsequently, total daily step counts were recorded either directly from the display of the devices (Yamax, Accupedo and Pedometer), or from the corresponding softwares after syncing (Actigraph, Vivofit and Vifit). All days, at which participants wore the devices simultaneously, were included in the analyses, regardless of the total daily wear time.

### 2.7. Statistical Analysis

The statistical analysis followed the validation and reporting standards developed by Welk et al. [33] and Johnston et al. [27]. Adherence to these standards ensured methodological and reporting consistency, facilitating comparison between wearable monitors and apps.

To facilitate comparison between devices and testing conditions and provide an indicator of overall measurement error, MAPE was used. A smaller MAPE represents better accuracy. Johnston et al. [27] recommend  $MAPE \leq 5\%$ , if the activity monitor is to be used as an outcome measure within a clinical trial or as an alternative gold-standard measurement tool for step counting, and  $MAPE \leq 10\text{--}15\%$  if the device is being validated for use by the general population.

To evaluate the level of agreement, Bland–Altman plots with corresponding 95% limits of agreement and fitted lines (from regression analyses between mean and difference) with their corresponding parameters (i.e., intercept and slope) were presented. A fitted line that provides a slope of 0 and an intercept of 0 exemplifies perfect agreement, while a statistically significant slope suggests that there is proportional systematic bias. Bland–Altman analysis is widely accepted as the most appropriate tool in assessing agreement within medical validation studies, providing a measure of the agreement between the two measurements [34–36].

Finally, for data collected in the field study, the within-device precision (reliability) of the devices was assessed by calculating the Intraclass Correlation Coefficient (ICC; two-way random, absolute agreement) and corresponding 95% confidence intervals (CI). The calculated ICC was used as the basis to assess the degree of agreement using the following guideline:  $<0.50$  poor;  $0.50$  to  $0.75$  moderate;  $0.75$  to  $0.90$  good; and  $>0.90$  excellent correlation [37]. ICC were not estimated for data collected in the lab-based study because the criterion steps was a constant variable (i.e., 400 steps), so the variance was equal to zero and any correlations could not be defined. The statistical analyses were performed with SPSS version 27.0 for Windows (IBM SPSS Corp., Armonk, NY, USA) and MedCalc 12.7 (MedCalc Software bvba).

## 3. Results

### 3.1. Lab-Based Semi-Structured Study

During the lab-based study, participants averaged 400 steps in total. The MAPE was low for all wearable activity monitors, ranging from 2.57% for Vivofit, up to 5.76% for Vifit. On the other hand, the magnitude of error for the two apps was significantly higher than 20%.

The Bland–Altman plots are included in the Supplementary file (Figures S1–S5). For walking, the plots revealed the narrowest 95% limits of agreement for Vivofit (difference =  $-3.21$  steps), while values were higher for Vifit (difference =  $-16.04$  steps) and Yamax (difference =  $16.75$  steps). The plots for both apps showed higher values of disagreement, over 100 steps. The slopes for all monitors and apps were significant, revealing the systematic overestimation of steps as measured by the Vivofit, Vifit, Accupedo

and Pedometer, and underestimation of Yamax measured steps, compared to the criterion measurement (direct observation). All results are presented in Table 2.

**Table 2.** Step count MAPE and Bland–Altman results for the lab-based semi-structured study (400 steps).

Device	M (SD)	MAPE		Bland–Altman			
		% Error	95% CI	M Diff	95% CI	Slope	<i>p</i>
Yamax	383 (21)	5.04	3.18–6.90	16.75	7.59–25.91	−2.11	<0.001
Vivofit	403 (13)	2.57	1.64–3.51	−3.21	(−9.09)–2.68	−2.12	<0.001
Vifit	416 (25)	5.76	3.82–7.70	−16.04	(−26.57)–(−5.51)	−1.98	<0.001
Accupedo	508 (58)	27.31	21.36–33.27	−108.33	(−133.00)–(−83.66)	−2.01	<0.001
Pedometer	563 (120)	41.02	28.57–53.47	−162.75	(−213.43)–(−112.07)	−2.01	<0.001

Note. CI: confidence interval; Diff: difference.

### 3.2. Field Study

During the 3-day field study, participants averaged  $5657 \pm 1751$  steps (as estimated by Actigraph). The MAPE was significantly higher for all monitors and apps (except one) compared to the field-based study. More specifically, Yamax and Vifit had the lowest magnitude of errors (12.85% and 12.87%, respectively), while Accupedo app had the highest magnitude of error (32.20%). The Pedometer app had significantly lower MAPE compared to the field-based study (13.59% vs. 41.02%).

The Bland–Altman plots are included in the Supplementary file (Figures S6–S10). The plots revealed the narrowest 95% limits of agreement for Yamax (difference =  $-71.11$  steps), while values were significantly higher for Vifit (difference =  $-696.21$  steps) and Vivofit (difference =  $-934.79$  steps). The plot for Accupedo showed a higher value of disagreement (almost 2000 steps), while the Pedometer app yielded a lower value than Vifit and Vivofit (difference =  $-582.68$  steps). The slopes for Accupedo, Vivofit, and Vifit were significant, revealing the systematic overestimation of steps compared to Actigraph results, while the slopes for Yamax and Pedometer did not yield any proportional systematic bias.

Lastly, the ICC was moderate to high for all monitors and apps. More specifically, Accupedo showed moderate correlation, Vifit, Vivofit, and Pedometer yielded good correlation, while Yamax showed excellent correlation with the criterion (i.e., Actigraph). All results are presented in Table 3.

**Table 3.** Step count MAPE, Bland–Altman and ICC results for the field study.

Device	M (SD)	MAPE			Bland–Altman				ICC	
		% Error	95% CI	M Diff	95% CI	Slope	<i>p</i>	Score	95% CI	
Yamax	5728 (2097)	12.85	8.87–16.83	−71.11	(−455.35)–313.14	−0.19	0.058	0.92	0.80–0.97	
Vivofit	6592 (2153)	16.73	11.60–21.85	−934.79	(−1271.79)–(−597.79)	−0.21	0.009	0.85	0.64–0.96	
Vifit	6353 (2142)	12.87	8.29–17.27	−696.21	(−1061.88)–(−330.54)	−0.21	0.022	0.87	0.42–0.96	
Accupedo	7509 (2484)	32.20	26.93–37.46	−1851.32	(−2308.04)–(−1394.59)	−0.35	<0.001	0.66	0.08–0.91	
Pedometer	6240 (2086)	13.59	7.73–19.65	−582.68	(−1004.74)–(−160.63)	−0.18	0.098	0.86	0.57–0.95	

Note. CI: confidence interval; Diff: difference.

## 4. Discussion

The aim of the present study was to examine the validity of three wearable monitors and two Android PA apps for measuring steps during semi-structured and free-living studies, in a sample of healthy adults. The results revealed high validity for the three wearable monitors during the semi-structured study in the lab, with MAPE values approximately 5% for Yamax and Vifit and well below 5% for Vivofit. This finding is in accordance with

previous literature, where it has been shown that wearable monitors usually achieve high levels of steps' criterion validity [13,14].

On the other hand, the two smartphone apps showed high MAPE values over 20%, overestimating by more than 100 steps compared to direct observation. Previous studies showed conflicting evidence on apps' validity to count steps [20]. For example, Adamakis [28] found that all freeware accelerometer-based apps were valid in all conditions that were tested, while Orr et al. [38] concluded that on the 20-step test none of the applications met the 5% error threshold. To explain this inconsistency, it is important to notice that the validation studies for step count which found that PA apps were likely to meet acceptable accuracy levels showed an increased accuracy at higher speeds [15,28,39]. Taking into consideration that during the semi-structured condition of the present study the participants walked at a slow-to-average speed, the increased apps' MAPE can potentially be attributed to this specific factor.

During the free-living study all monitors and apps, even though correlated moderate-to-high with the criterion measurement (i.e., Actigraph), had high MAPE, over 10%. The lowest error was observed for Yamax, Vifit and Pedometer app, while Accupedo app had the highest error, overestimating steps by 32%. It is not surprising that higher measurement errors were found in free-living conditions than semi-structured lab settings. Several studies that have been conducted in both settings have come to a similar conclusion [14]. Bai et al.'s [20] study revealed low to acceptable validity from three wearable monitors in free-living settings in estimating steps, with an overall error in steps of 20%. Similarly, Duncan et al. [40] found that during lab tests, criterion and iPhones differed from manually counted steps by a mean bias of less than 5%, while in the free-living condition steps differed by a mean bias of 21.5% or 1340 steps/day.

In general, it seems that wearable monitors and smartphone PA apps measure steps more accurately in controlled or semi-structured settings than free-living ones [14,15,18–20,38,40,41]. Usually, the errors in the free-living settings tend to be higher than 10% [15], which has been confirmed by the findings of the present study. This low result is crucial for monitoring steps during daily activities, mainly because under free-living conditions, where intervention studies require the highest validity, the wearable monitors and apps are not valid, and commonly underreport/overreport steps.

Based on Johnston et al.'s [27] recommendations, the three wearable monitors under examination have the potential to be used as step outcome measures within clinical trials or as alternative gold-standard measurement tools for step counting only in semi-structured settings, since their MAPE is lower than 5%. On the other hand, apps in both settings and monitors in free-living settings cannot be considered valid instruments for measuring steps. Individuals who primarily walk and perform light, intermittent lifestyle activities such as the sample of the present study, as well as researchers (especially those who conduct large-scale epidemiological studies) should be cautioned in considering the use of smartphone apps and wearable monitors as research grade monitors for PA surveillance or evaluation using steps as an outcome measure. Of course, more validation studies should be carried out to further support or contradict these findings.

Collectively, this study and previous works cannot support the value of wearable monitors and apps as acceptable measures of PA and step count in free-living contexts. Even though further value is added by passively measuring PA in large population groups without the use of dedicated measurement tools, as well as by the capacity to generate large datasets that could be used to understand temporal, location, and contextual factors that affect PA, caution should be taken when these devices are used for research purposes. Certain limitations regarding their validity and reliability should always be acknowledged and taken into consideration.

This study is not without limitations. Participants included in the current study were healthy adults in their early 30 s, and the sample size was limited. Additional research is needed to assess the validity of the monitors and apps in other special populations, particularly in those without a typical locomotive pattern. Future studies should include

participants of non-stereotypical gait, older individuals, individuals from different ethnic groups, and much larger heterogeneous groups. Another limitation is the use of two Android apps; it is not clear whether our results are applicable to other smartphone apps and operating systems (i.e., iOS). A third limitation of the study was the fact that video recording was not used as the criterion measure during the free-living field study. While Actigraph has proven validity to record steps in free-living settings, it counts steps based on accelerometry and thus may suffer from inaccuracies at lower speeds. The use of video recording could have improved the accuracy of the true step counts; however, due to the 3-day monitoring period, it was deemed the use of Actigraph more feasible and less likely to affect participants' compliance. Finally, the role of the smartphone's optimal position on the human body during exercise should be further investigated.

## 5. Conclusions

PA tracking monitors and freeware apps have the potential to capture real-time step data and are used in large cohort studies to estimate how PA is associated with mortality risk. However, the validity of numerous commercially available apps and monitors, especially in free-living settings, remains unclear. In this validation study, the results suggested that the three wearable monitors under examination (i.e., Yamax, Vivofit and Vifit) were valid in the semi-structured lab-based context and could be considered suitable for use as step outcome measures within a clinical trial. On the other hand, these monitors were not valid in free-living settings, showing high systematic errors. Wearable monitors that might be valid in one context, might not be valid in different contexts and vice versa, and researchers should be aware of this specific monitors' limitation. The apps under examination (i.e., Accupedo and Pedometer) were not valid in both conditions, showing high MAPE over 10%. Caution is required in relying on these apps for outcome measures of PA within intervention trials and observational studies. In addition, given the importance of self-monitoring for behavior change, care is required in the promotion of these apps for use by the public. As companies develop and release new wearable monitors and smartphone apps, usually they do not make available the method for calculating steps, so researchers will have to continue examining the accuracy and validity of these devices to provide accurate information to consumers and researchers.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/technologies11010029/s1>, Figures S1–S10: All Bland–Altman plots for step count comparisons with criterion measure (10 plots in total).

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Social Research Ethics Committee of UNIVERSITY COLLEGE CORK in Ireland.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to reasons of sensitivity and restrictions, as these contain information that could compromise research participants' privacy/consent.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Physical Activity Guidelines Advisory Committee. *Physical Activity Guidelines Advisory Committee Scientific Report*; Department of Health and Human Services: Washington, DC, USA, 2018. Available online: [https://health.gov/sites/default/files/2019-09/PAG\\_Advisory\\_Committee\\_Report.pdf](https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf) (accessed on 15 July 2022).
2. Saint-Maurice, P.F.; Troiano, R.P.; Bassett, D.R.; Graubard, B.I.; Carlson, S.A.; Shiroma, E.J.; Fulton, J.E.; Matthews, C.E. Association of daily step count and step intensity with mortality among US adults. *JAMA* **2020**, *323*, 1151–1160. [CrossRef] [PubMed]

3. Paluch, A.E.; Bajpai, S.; Bassett, D.R.; Carnethon, M.R.; Ekelund, U.; Evenson, K.R.; Galuska, D.A.; Jefferis, B.J.; Kraus, W.E.; Lee, I.M.; et al. Daily steps and all-cause mortality: A meta-analysis of 15 international cohorts. *Lancet Public Health* **2022**, *7*, e219–e228. [[CrossRef](#)] [[PubMed](#)]
4. Paluch, A.E.; Gabriel, K.P.; Fulton, J.E.; Lewis, C.E.; Schreiner, P.J.; Sternfeld, B.; Sidney, S.; Siddique, J.; Whitaker, K.M.; Carnethon, M.R. Steps per day and all-cause mortality in middle-aged adults in the coronary artery risk development in young adults study. *JAMA Netw. Open* **2021**, *4*, e2124516. [[CrossRef](#)] [[PubMed](#)]
5. Hall, K.S.; Hyde, E.T.; Bassett, D.R.; Carlson, S.A.; Carnethon, M.R.; Ekelund, U.; Evenson, K.R.; Galuska, D.A.; Kraus, W.E.; Lee, I.M.; et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. *Int. J. Behav. Nutr. Phys. Act.* **2020**, *17*, 78. [[CrossRef](#)]
6. Thompson, W. Worldwide survey of fitness trends for 2020. *ACSM's Health Fit. J.* **2019**, *23*, 10–18. [[CrossRef](#)]
7. Thompson, W. Worldwide survey of fitness trends for 2022. *ACSM's Health Fit. J.* **2022**, *26*, 11–20. [[CrossRef](#)]
8. Thompson, W. Worldwide Survey of Fitness Trends for 2021. *ACSM's Health Fit. J.* **2021**, *25*, 10–19. [[CrossRef](#)]
9. Laricchia, F. Total Wearable Device Unit Shipments Worldwide 2014–2021. 2022. Available online: <https://www.statista.com/statistics/437871/wearables-worldwide-shipments/> (accessed on 10 November 2022).
10. Laricchia, F. Global Smartphone Sales to End Users 2007–2021. 2022. Available online: <https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/> (accessed on 8 November 2022).
11. Laranjo, L.; Ding, D.; Heleno, B.; Kocaballi, B.; Quiroz, J.C.; Tong, H.L.; Chahwan, B.; Neves, A.L.; Gabarron, E.; Dao, K.P.; et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *Br. J. Sports Med.* **2021**, *55*, 422–432. [[CrossRef](#)]
12. Nelson, M.B.; Kaminsky, L.A.; Dickin, D.C.; Montoye, A.H. Validity of consumer-based physical activity monitors for specific activity types. *Med. Sci. Sports Exerc.* **2016**, *48*, 1619–1628. [[CrossRef](#)]
13. Dowd, K.P.; Szeklicki, R.; Minetto, M.A.; Murphy, M.H.; Polito, A.; Ghigo, E.; van der Ploeg, H.; Ekelund, U.; Maciaszek, J.; Stemplewski, R.; et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: A DEDIPAC study. *Int. J. Behav. Nutr. Phys. Act.* **2018**, *8*, 15. [[CrossRef](#)]
14. Fuller, D.; Colwell, E.; Low, J.; Orychock, K.; Tobin, M.A.; Simango, B.; Buote, R.; Van Heerden, D.; Luan, H.; Cullen, K.; et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth uHealth* **2020**, *8*, e18694. [[CrossRef](#)]
15. Silva, A.G.; Simões, P.; Queirós, A.; Rodrigues, M.; Rocha, N.P. Mobile apps to quantify aspects of physical activity: A systematic review on its reliability and validity. *J. Med. Syst.* **2020**, *8*, 51. [[CrossRef](#)]
16. Ferguson, T.; Rowlands, A.V.; 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. [[CrossRef](#)]
17. Breteler, M.J.; Janssen, J.H.; Spiering, W.; Kalkman, C.J.; van Solinge, W.W.; Dohmen, D.A.J. Measuring free-living physical activity with three commercially available activity monitors for telemonitoring purposes: Validation study. *JMIR Form Res.* **2019**, *3*, e11489. [[CrossRef](#)]
18. Vetrovsky, T.; Siranec, M.; Marenčáková, J.; Tufano, J.J.; Capek, V.; Bunc, V.; Belohlavek, J. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. *PLoS ONE* **2019**, *14*, e0222569. [[CrossRef](#)]
19. Höchsmann, C.; Knaier, R.; Infanger, D.; Schmidt-Trucksäss, A. Validity of smartphones and activity trackers to measure steps in a free-living setting over three consecutive days. *Physiol. Meas.* **2020**, *41*, 015001. [[CrossRef](#)]
20. Bai, Y.; Tompkins, C.; Gell, N.; Dione, D.; Zhang, T.; Byun, W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. *PLoS ONE* **2021**, *5*, e0251975. [[CrossRef](#)]
21. Amagasa, S.; Kamada, M.; Sasai, H.; Fukushima, N.; Kikuchi, H.; Lee, I.M.; Inoue, S. How well iPhones measure xsteps in free-living conditions: Cross-sectional validation study. *JMIR mHealth uHealth* **2019**, *9*, e10418. [[CrossRef](#)]
22. Shei, R.J.; Holder, I.G.; Oumsang, A.S.; Paris, B.A.; Paris, H.L. Wearable activity trackers—advanced technology or advanced marketing? *Eur. J. Appl. Physiol.* **2022**, *122*, 1975–1990. [[CrossRef](#)]
23. Hartung, V.; Sarshar, M.; Karle, V.; Shammas, L.; Rashid, A.; Roullier, P.; Eilers, C.; Mäurer, M.; Flachenecker, P.; Pfeifer, K.; et al. Validity of consumer activity monitors and an algorithm using smartphone data for measuring steps during different activity types. *Int. J. Behav. Nutr. Phys. Act.* **2020**, *17*, 9314. [[CrossRef](#)]
24. Keadle, S.K.; Lyden, K.A.; Strath, S.J.; Staudenmayer, J.W.; Freedson, P.S. A framework to evaluate devices that assess physical behavior. *Exerc. Sport. Sci. Rev.* **2019**, *47*, 206–214. [[CrossRef](#)] [[PubMed](#)]
25. Zou, G.Y. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat. Med.* **2012**, *31*, 3972–3981. [[CrossRef](#)] [[PubMed](#)]
26. Evenson, K.R.; Goto, M.M.; Furberg, R.D. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int. J. Behav. Nutr. Phys. Act.* **2015**, *12*, 159. [[CrossRef](#)] [[PubMed](#)]
27. Johnston, W.; Judice, P.B.; García, P.M.; Mühlén, J.M.; Skovgaard, E.L.; Stang, J.; Schumann, M.; Cheng, S.; Bloch, W.; Brønd, J.C.; et al. Recommendations for determining the validity of consumer wearable and smartphone step count: Expert statement and checklist of the INTERLIVE network. *Br. J. Sports Med.* **2021**, *55*, 780–793. [[CrossRef](#)] [[PubMed](#)]
28. Adamakis, M. Criterion validity of iOS and Android applications to measure steps and distance in adults. *Technologies* **2021**, *9*, 55. [[CrossRef](#)]

29. Adamakis, M. Validity of wearable monitors and smartphone applications to measure steps and distance in adolescents. *Sport Mont* **2022**, *20*, 3–10. [[CrossRef](#)]
30. Huang, Y.; Xu, J.; Yu, B.; Shull, P.B. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait Posture* **2016**, *48*, 36–41. [[CrossRef](#)]
31. Sasaki, J.E.; John, D.; Freedson, P.S. Validation and comparison of ActiGraph activity monitors. *J. Sci. Med. Sport* **2011**, *14*, 411–416. [[CrossRef](#)]
32. John, D.; Freedson, P. ActiGraph and Actical physical activity monitors: A peek under the hood. *Med. Sci. Sports Exerc.* **2012**, *44*, S86–S89. [[CrossRef](#)]
33. Welk, G.J.; Bai, Y.; Lee, J.M.; Godino, J.; Saint-Maurice, P.F.; Carr, L. Standardizing analytic methods and reporting in activity monitor validation studies. *Med. Sci. Sports Exerc.* **2019**, *51*, 1767–1780. [[CrossRef](#)]
34. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *327*, 307–310. [[PubMed](#)]
35. Zaki, R.; Bulgiba, A.; Ismail, R.; Ismail, N.A. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. *PLoS ONE* **2012**, *7*, e37908. [[CrossRef](#)]
36. Ludbrook, J. Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clin. Exp. Pharmacol. Physiol.* **2002**, *29*, 527–536. [[CrossRef](#)]
37. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)]
38. Orr, K.; Howe, H.S.; Omran, J.; Smith, K.A.; Palmateer, T.M.; Ma, A.E.; Faulkner, G. Validity of smartphone pedometer applications. *BMC Res. Notes* **2015**, *8*, 733. [[CrossRef](#)]
39. Höchsmann, C.; Knaier, R.; Eymann, J.; Hintermann, J.; Infanger, D.; Schmidt-Trucksäss, A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. *Scand. J. Med. Sci. Sports* **2018**, *28*, 1818–1827. [[CrossRef](#)]
40. Duncan, M.J.; Wunderlich, K.; Zhao, Y.; Faulkner, G. Walk this way: Validity evidence of iphone health application step count in laboratory and free-living conditions. *J. Sports Sci.* **2018**, *36*, 1695–1704. [[CrossRef](#)]
41. Leong, J.Y.; Wong, J.E. Accuracy of three Android-based pedometer applications in laboratory and free-living settings. *J. Sports Sci.* **2017**, *35*, 14–21. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.