DOI 10.26773/smj.220601

ORIGINAL SCIENTIFIC PAPER

Validity of Wearable Monitors and Smartphone Applications to Measure Steps and Distance in Adolescents

Manolis Adamakis[1,2]

[1]University College Cork, School of Education, Cork, Ireland, [2]Department of Education and Social Work, University of Luxembourg, Luxembourg

Abstract

The growing popularity of wearable physical activity (PA) monitors and fitness applications (apps) in recent years and the vast amounts of data that they generate present attractive possibilities for surveillance. However, measurement accuracy is indispensable when tracking PA variables to provide meaningful measures of PA. The purpose of this study was to examine the criterion validity of wearable PA monitors and a combination of GPS and accelerometer free of charge smartphone apps, during self-paced outdoor walking and running. Thirty-eight healthy adolescents (15.3±2.0 years) participated in this cross-sectional study. They were fitted with Garmin Forerunner 310XT, Garmin Vivofit, Medisana Vifit, and smartphones running the Runkeeper, Runtastic, Sports Tracker (GPS), Pedometer, Accupedo, Pedometer and Pedometer 2.0 (accelerometer) apps. They were asked to walk and run 1.22 km for each trial and two researchers counted every step taken during trials with a digital tally counter. Validity was evaluated by comparing each device with the criterion measure using Repeated measures analysis of variance (ANOVA), Mean Absolute Percentage Errors (MAPE) and Bland-Altman plots. MAPE were low for Forerunner and GPS apps for distance in both conditions (2.27%-9.73%), and significantly higher for the accelerometer monitors and apps (6.92%-39.02%). Vivofit (MAPE=6.51%) and Vifit (MAPE=6.66%) accurately estimated the number of steps during walking, however only Vivofit (MAPE=3.95%) was accurate during running. All accelerometer-based apps had high MAPE for step counting (9.87%-40.26%). The findings suggested that GPS monitors and apps were accurate tools for counting distance during walking and running, while accelerometer-based monitors and apps had higher errors. Vivofit provided accurate estimates of step count in both conditions, and Medisana Vifit was valid during walking. Accupedo was the only app with an acceptable step count error.

Keywords: *wearable activity tracker, GPS, accelerometer, fitness tracker, step count*

Introduction

The World Health Organization's (WHO) guidelines on physical activity (PA) and sedentary behavior were recently updated (WHO, 2020). Children and adolescents should complete at least an "average of 60 minutes/day" of moderate-to-vigorous physical activity (MVPA), rather than the previously stated "60 minutes/day of PA" as indicated in 2010 WHO recommendations (Bull et al., 2020). Updated PA guidelines pose challenges to PA surveillance; however, device-based measurements may facilitate surveillance (Troiano, Stamatakis, & Bull, 2020).

Wearable PA monitors and fitness applications (apps) have grown in popularity in recent years, and the vast volumes of data they create present appealing surveillance opportunities (Omura, Carlson, Paul, Watson, & Fulton, 2017). Unfortunately, these consumer-targeted wearable technologies designed for general wellness purposes are not required to go through a standardized evaluation process to ensure their accuracy, validity and reliability, and most product manufacturers only provide the minimum requirements when releasing information (Bent & Dunn, 2020). Nevertheless, for any research project where PA

Correspondence:

M. Adamakis
School of Education, University College Cork, Sports Studies and Physical Education, 2 Lucan place, Cork, Ireland
E-mail: manosadam@phed.uoa.gr

is either an end measure or an intervention, accurate and reliable measurements are crucial for tracking PA variables in order to produce relevant PA estimates (Nelson, Kaminsky, Dickin, & Montoye, 2016).

Most previous studies on the validity of PA monitors and smartphone apps were conducted on adults. According to a recent review of step counting wearable technologies' validation on a treadmill, Mean Absolute Percentage Error (MAPE) values were 7% to 11% for wrist-worn, 1% to 4% for waist-worn, and ≤1% for thigh-worn monitors (Moore, McCullough, Aguiar, Ducharme, & Tudor-Locke, 2020). Two other systematic reviews examined the validity of Fitbit and Garmin monitors. Fitbit devices were likely to meet acceptable accuracy for step count approximately half the time, with a tendency to underestimate steps in controlled testing and overestimate steps in free-living settings (Feehan et al., 2018). Step counting accuracy was likewise high on Garmin activity monitors, however distance validity was low, with MAPE surpassing the acceptable limits (Evenson & Spade, 2020). Further review of the literature suggests that all Global Positioning System (GPS) units, regardless of sampling rate, are valid for tracking distance in team sports, with adequate intra-unit reliability to allow multiple comparisons of a single device (Scott, Scott, & Kelly, 2016).

Regarding the validity of smartphone apps, a recent review concluded that there is conflicting and insufficient evidence on the validity and reliability of apps for measuring PA. Nevertheless, velocity and the place where the smartphone is carried seem to have an impact on validity, as absolute errors decreased with higher velocities (Silva, Simões, Queirós, Rodrigues, & Rocha, 2020). In this review, only two studies tested the apps for distance validity in outdoor settings. The first study tested the app during walking and running in a circuit with a known distance and with the app attached to the arm (Adamakis, 2017) and the second study during a marathon/half-marathon with no indication where the app was carried (Pobiruchin, Suleder, Zowalla, & Wiesner, 2017). Apps' MAPE in both studies were less than 4.5% (Adamakis, 2017; Pobiruchin, Suleder, Zowalla, & Wiesner, 2017).

Due to the fact that there is an apparent potential of PA monitors and apps to measure and promote PA, there is a need to carry out more studies of high methodological quality in various populations. Thus, the purpose of the present study was to validate step count and distance travelled of one GPS and two consumer-based monitors, as well as six Android apps (three pedometer- and three GPS-based apps) in a sample of healthy adolescents. Based on the evaluation framework proposed by Keadle, Lyden, Strath, Staudenmayer and Freedson (2019), a naturalistic validation study design in real-world conditions was used, which included self-paced outdoor walking and running. The submaximal outdoor walking and running tests were performed in regular outdoor conditions with the aim of providing data from uncontrolled and sometimes challenging conditions, where participants would train and perform their regular fitness activities.

Methods

Study design

This was a cross-sectional study investigating the accuracy in recorded distance and step count values for nine wearable technology monitors and apps. The research design was similar to a previously published validation study (Adamakis, 2020). Thirty-eight healthy adolescents, with no contraindications for exercise and no known orthopaedic limitations that would prevent them from completing the assessments, participated. All adolescents, as well as their parents, read and signed an informed consent form approved by the School of Sport Science and Physical Education of Athens Research Ethics Committee, informing them of the risks and benefits of the study.

Participants reported to the researchers twice. During the first visit, anthropometric measures were obtained in controlled laboratory settings. The second visit (2 - 3 days after the first visit) took place in a track and field elliptical stadium. Field tests were performed outdoors between November and January, during days when it was not raining, and the temperature was above 10°C. These conditions are typical outdoor training conditions and, hence, provide a good benchmark for challenging real outdoor training conditions that are faced by adolescents while exercising. All participants were instructed to wear their own outdoor sports clothing as appropriate for the current weather during the test.

The participants were fitted with three different activity trackers and three smartphones, each one running simultaneously two different apps (one GPS and one pedometer-based app). Garmin Vivofit (Vivofit) and Garmin Forerunner 310XT (Forerunner) were both worn on the left wrist. Medisana Vifit (Vifit), as well as the three smartphones, 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 addition to the devices, a heart rate monitor (Garmin HRM-Dual™) was placed around participants' chest to capture exercise heart rate.

Participants had to perform a total of two field tests in regular outdoor conditions: overground walking and submaximal running, at a self-selected pace. The only limitation that existed was that walking speed should be between 3 and 6 km/h and running speed should be above 8 km/h, following the American College of Sports Medicine (2006) recommendations (speed between 6 and 8 km/h is considered a transitional speed between walking and running and should be avoided in experimental procedures). The actual average walking speed, estimated by Forerunnner, was 5.27±0.62 km/h (1.47±0.17 m/s) for walking and 11.05±1.47 km/h (3.07±0.41 m/s) for running, respectively. Between the two trials, participants could rest for 5 minutes and all devices were paused simultaneously. During pause, all apps' specific settings were changed from walking to running option.

Distance was objectively recorded with a manual distance measuring wheel [Roadrunner RR182 (Keson)], by measuring the walking route two times and then taking the mean distance for an ending point. The total distance that all participants had to walk, and run was 1.22 km for each trial (2.44 km in total). In addition, two manual counters objectively measured steps. For all self-paced walks and runs they were instructed to follow the participants and were separated so they could not view each other's thumb motion nor hear the "clicking" from the tally counter. This prevented any synchronized counting between the two. The reliability of this method was tested by comparing a video recording for two walking and running video sequences of one of the subjects. An intra-class correlation coefficient value of 0.99 was obtained through the analyses of the video sequences and the steps recorded by the researchers. Smartphones were set to airplane mode to avoid interactions with the mobile phone providers (i.e., no data connection), and all devices were activated simultaneously. In the end of each trial, data initially were stored manually and at a later time were uploaded to the relating devices' software.

Participants

A power calculation with findings of observed step counts (correlation of 0.50), alpha two-tailed value of 0.05, and a power of 0.80 indicated a sample size of 29 participants. In total, 38 healthy adolescents (n=16 boys, n=22 girls) with an age range of 12-18 years (15.7±1.8 years), body mass index range of 15.1 - 28.6 kg/m2 (21.1±2.3 kg/m^2) were screened and participated in the study (with no dropouts).

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 was calculated as weight (kg) / height squared (m^2). 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±SD)

	Boys	Girls	Total
	M±SD	M±SD	M±SD
Age (years)	15.3±2.0	16.0±1.7	15.7±1.8
Weight (kg)	70.1±11.4	55.4±6.5	61.6±8.2
Height (m)	1.76±0.09	1.66±0.05	1.70±0.09
Body Mass Index (kg/m^2)	22.4±3.6	20.1±1.8	21.1±2.3
Resting heart rate (bpm)	71.4±9.3	67.4±7.4	69.1±8.4
Walking step length (cm)	79.6±5.8	77.7±7.1	78.5±6.6
Running step length (cm)	124.6±20.8	120.2±16.7	122.1±18.4

Wearable monitors

Garmin Forerunner 310XT (Forerunner): Forerunner (Garmin International Inc., Olathe, KS, U.S.A.) is a mid-cost GPS-enabled training and heart rate wrist-worn monitor for multisport athletes. It tracks time, distance, average and lap speed and pace, heart rate with a premium heart rate monitor, on land and estimated calorie burn.

Garmin Vivofit (Vivofit): Vivofit (Garmin International Inc., Olathe, KS, U.S.A.) is a mid-cost wrist-based, triaxial accelerometer-based monitor that measures steps taken, distance travelled, calories expended and sleep quality. The Garmin Connect software was used to upload data for both Forerunner and Vivofit.

Medisana Vifit (Vifit): Vifit (Medisana AG, Neuss, Germany) 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 PA 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 and distance data.

Android apps

This study used three Samsung smartphones S8 based on the Android operating system. Inclusion criteria for all apps were retrieved from previous protocols (Adamakis, 2020, 2021): (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 for the pedometers and GPS for the GPS apps (no 4G/5G signal); (4) Ability to record the number of steps taken, average speed, total distance and energy expenditure; (5) Adjustable sensitivity settings for the pedometers; (6) Manual input of demographic and somatometric data (sex, age, weight, height and step length for walking and running) for accurate EE estimation; (7) Manual choice of activity type (i.e. walking or running); (8) Among the most popular and downloadable applications, according to users' ratings and number of downloads from the Google Play Store. Specifically, for the pedometer apps, they should include an option to capture step count during walking and running separately, by inputting different step lengths for the two conditions.

Based on the previously described criteria, three GPS- and three accelerometer-based apps were selected: (1. GPS) Runkeeper (ASICS Digital, Inc.), Runtastic (Runtastic GmbH), Sports Tracker (Sports Tracking Technologies); (2. Accelerometer) Pedometer (ITO Technologies, Inc.), Accupedo (Corusen LLC), Pedometer 2.0 (DSD). Pedometer 2.0 was the only application with a self-calibration capability, which was used to determine the appropriate sensitivity settings for every participant separately. All wearable monitors and apps were included in the validation of distance travelled, however only pedometer monitors (i.e., Vivofit and Vifit) and apps (i.e., Pedometer, Accupedo, and Pedometer 2.0) were tested for step count validity.

Statistical analysis

Descriptive analyses were conducted to examine associations with the criterion measure. Repeated measures analysis of variance (ANOVA) statistical tests were performed to assess differences from all monitors and apps, and criterion measures for distance and step count. When the test statistic was significant, post-hoc pairwise comparisons with Bonferroni correction were performed. The significance level was set at $p<0.05$ and the partial η^2 was presented as a measure of effect size for F-tests. A partial η^2 value between 0.01 and 0.06 was associated with a small effect, between 0.06 and 0.14 with a medium effect, and 0.14 or greater with a large effect (Warner, 2012).

MAPE were also calculated to provide an indicator of overall measurement error {MAPE = [(monitor measurement-criterion measure) / criterion measure] x 100} and was used as an outcome measure. A smaller MAPE represents better accuracy. Johnston et al. (2020) recommend MAPE ≤5%, if the PA 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 ≤10%-15% if the device is being validated for use by the general population.

To further evaluate individual variations, 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 (Bland & Altman, 1986; Ludbrook, 2002). A fitted line that provides a slope of 0 and an intercept of 0 exemplifies perfect agreement. 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).

Results

Step count

Participants averaged 1450±119 steps during walking and 1070±134 steps during running trials, respectively. The repeated measures ANOVA for both walking [$F(2,74)=29.30$, $p<0.001$, $\eta^2=0.44$] and running [$F(2,74)=7.25$, $p=0.001$, $\eta^2=0.16$] were statistically significant, with large effect sizes. The post-hoc pairwise comparisons with Bonferroni corrections showed that only Vivofit [$F(1,37)=1.95$, $p=0.170$] did not differ significantly from the criterion during walking, while Vivofit [$F(1,37)=0.46$, $p=0.500$] and Accupedo [$F(1,37)=0.79$, $p=0.380$] did not differ significantly from the criterion during running. All remaining comparisons with the criterion resulted in statistically significant differences ($p<0.05$) (Table 2).

Table 2. Results of repeated measures ANOVA for step count and comparison with criterion measure

	Walking					Running				
	M	SD	Pairwise F	Pairwise p	95% CI	M	SD	Pairwise F	Pairwise p	95% CI
Criterion	1450	119	-	-	-	1070	134	-	-	-
Vivofit	1417	193	1.95	0.17	(-42) - 109	1078	120	0.46	0.50	(-49) - 31
Vifit	1532	110	31.89	<0.001	(-127) - (-36)	942	190	31.84	<0.001	57 - 199
Pedometer	1958	495	39.13	<0.001	(-762) - (-253)	1233	344	5.68	0.02	(-378) - 52
Accupedo	1537	146	11.17	0.002	(-168) - (-5)	1104	245	0.79	0.38	(-157) - 88
Pedometer 2.0	1973	493	41.74	<0.001	(-777) - (-269)	1217	375	4.40	0.04	(-367) - 73

Note. CI: confidence interval

During walking MAPE was least for monitors, Vivofit (6.51%) and Vifit (6.66%), while error rates for apps were higher (Accupedo: 9.87%; Pedometer: 39.43%; Pedometer 2.0: 40.26%). During running the magnitude of errors was least for Vivofit (3.95%). Error rates for the remaining monitors and apps ranged from 11.69% to 34.02% (Figure 1).

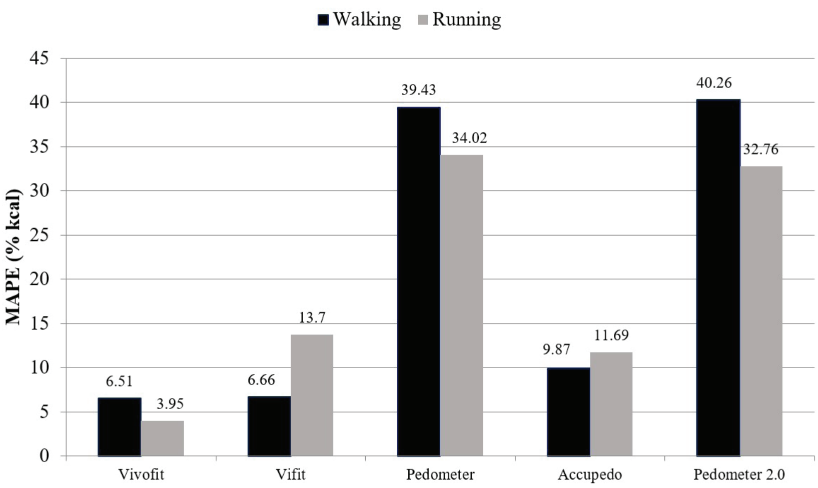

FIGURE 1. MAPE (% steps) of PA monitors and apps compared with criterion measure

The Bland-Altman results for step count for walking and running trials are presented in Table 3. For walking, the plots revealed the narrowest 95% limits of agreement for Vivofit (difference=33 steps), while values were the highest for Pedometer (difference=-507 steps) and Pedometer 2.0 (difference=-523 steps). During running, the narrowest 95% limits of agreement were observed for Vivofit (difference=-9 steps), followed by Accupedo (difference=-35 steps). The highest values were observed for Pedometer 2.0 (difference=-147 steps) and Pedometer (difference=-163 steps).

Distance

Participants averaged 1.22±0.01 km for both walking and running trials, respectively. The repeated measures ANOVA for both walking [$F(3, 104)=15.79$, $p<0.001$, $\eta^2=0.30$] and running [$F(3,104)=32.80$, $p<0.001$, $\eta^2=0.47$] were statistically significant, with large effect sizes. The post-hoc pairwise comparisons with Bonferroni corrections showed that only

Table 3. Step count Bland-Altman results during walking and running

	Walking					Running				
	M diff	95% CI	Slope	p	95% CI	M diff	95% CI	Slope	p	95% CI
Vivofit	33.45	(-15.07) - 81.96	-0.57	<0.001	(-0.86) - (-0.28)	-8.68	(-34.61) - 17.24	0.14	0.19	(-0.07) - 0.35
Vifit	-81.47	(-110.71) - (-52.24)	0.09	0.52	(-0.19) - 0.37	127.76	81.89 - 173.64	-0.39	0.008	(-0.68) - (-0.11)
Pedometer	-507.34	(-671.67) - (-343.01)	-1.72	<0.001	(-2.02) - (-1.43)	-163.00	(-301.58) - (-24.42)	-2.08	<0.001	(-2.67) - (-1.49)
Accupedo	-86.45	(-138.86) - (-34.03)	-0.32	0.20	(-0.81) - 0.18	-34.66	(-113.92) - 44.60	-0.84	<0.001	(-1.28) - (-0.41)
Pedometer 2.0	-522.84	(-686.82) - (-358.86)	-1.73	<0.001	(-2.03) - (-1.43)	-147.16	(-289.25) - (-5.06)	-1.86	<0.001	(-2.37) - (-1.36)

Note. CI: confidence interval; diff: difference

Sports Tracker [$F(1,37)=0.48$, $p=0.500$], Vifit [$F(1,37)=3.37$, $p=0.070$], and Accupedo [$F(1,37)=0.40$, $p=0.530$] did not differ significantly from the criterion during walking, while Sports Tracker [$F(1,37)=2.27$, $p=0.140$] and Accupedo [$F(1,37)=2.63$, $p=0.110$] did not differ significantly from the criterion during running. All remaining comparisons with the criterion resulted in statistically significant differences ($p<0.05$) (Table 4).

Figure 2 reports MAPE for all monitors and apps. During walking the magnitude of errors was least for Forerunner (2.27%) and Sports Tracker (3.92%), while error rates for all other were above 6.00% (6.82% - 35.94%). During running the magnitude of errors was least for Sports Tracker (3.07%), followed by Forerunner (3.40%) and Runkeeper (4.27%). Error rates for the remaining monitors and apps ranged from 8.83% to 39.02%.

Table 4. Results of repeated measures ANOVA for distance (km) and comparison with criterion measure

	Walking					Running				
	M	SD	Pairwise F	Pairwise p	95% CI	M	SD	Pairwise F	Pairwise p	95% CI
Criterion	1.22	0.01	-	-	-	1.22	0.01	-	-	-
Forerunner	1.24	0.04	4.48	0.04	(-0.03) - 0.01	1.26	0.03	58.66	<0.001	(-0.06) - 0.02
Runkeeper	1.30	0.14	10.83	0.002	(-0.15) - 0.05	1.25	0.07	4.27	0.04	(-0.06) - 0.02
Runtastic	1.15	0.15	8.31	0.007	(-0.02) - 0.16	1.13	0.07	78.47	<0.001	0.06 - 0.15
Sports Tracker	1.24	0.10	0.48	0.50	(-0.07) - 0.05	1.24	0.07	2.27	0.14	(-0.06) - 0.02
Vivofit	1.09	0.18	21.80	<0.001	0.03 - 0.24	1.14	0.20	7.33	0.01	(-0.03) - 0.20
Vifit	1.20	0.10	3.37	0.07	(-0.03) - 0.08	0.75	0.17	314.27	<0.001	0.38 - 0.57
Pedometer	1.54	0.56	11.48	0.002	(-0.63) - 0.01	0.99	0.30	23.75	<0.001	0.07 - 0.41
Accupedo	1.25	0.21	0.40	0.53	(-0.14) - 0.10	1.31	0.32	2.63	0.11	(-0.27) - 0.10
Pedometer 2.0	1.62	0.49	25.42	<0.001	(-0.67) - (-0.12)	1.46	0.42	11.73	0.002	(-0.48) - 0.01

Note. CI: confidence interval

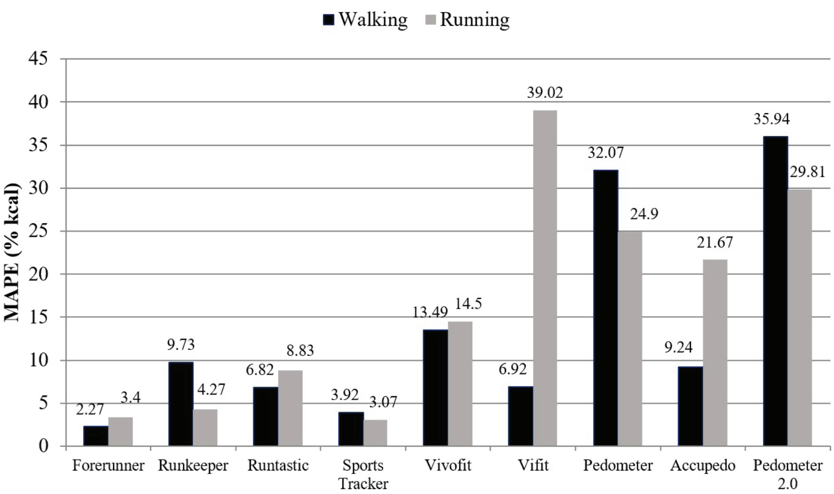

FIGURE 2. MAPE (% km) of PA monitors and apps compared with criterion measure

The Bland-Altman results for distance for both walking and running trials are presented in Table 5. For walking, the plots revealed the narrowest 95% limits of agreement for Forerunner (difference=-0.01 km) and Sports Tracker (difference=-0.01 km), and slightly higher values for Accupedo (difference=-0.02 km) and Vifit (difference=0.03 km), while values were the highest for Pedometer (difference=-0.31 km) and Pedometer 2.0 (difference=-0.40 km). During running, the narrowest 95% limits of agreement were observed for Sports Tracker (difference=-0.02 km), followed by Forerunner (difference=-0.04 km) and Runkeeper (difference=-0.05 km). The highest values were observed for Pedometer (difference=0.24 km) and Vifit (difference=0.48 km).

Table 5. Distance Bland-Altman results during walking and running

	Walking					Running				
	M diff	95% CI	Slope	p	95% CI	M diff	95% CI	Slope	p	95% CI
Forerunner	-0.01	(-0.02) - 0.00	-1.43	<0.001	(-1.72) - (-1.14)	-0.04	(-0.05) - (-0.03)	-1.35	<.001	(-1.66) - (-1.04)
Runkeeper	-0.07	(-0.12) - (-0.03)	-1.94	<0.001	(-2.05) - (-1.84)	-0.05	(-0.05 - 0.00	-2.03	<.001	(-2.25) - (-1.80)
Runtastic	0.07	0.02 - 0.12	-2.01	<0.001	(-2.11) - (-1.92)	0.11	0.08 - 0.13	-2.17	<.001	(-2.39) - (-1.95)
Sports Tracker	-0.01	(-0.04) - 0.02	-1.93	<0.001	(-2.08) - (-1.79)	-0.02	(-0.04) - 0.01	-1.83	<.001	(-2.02) - (-1.63)
Vivofit	0.14	0.08 - 0.18	-1.93	<0.001	(-2.00) - (-1.85)	0.09	0.02 - 0.15	-1.92	<.001	(-1.99) - (-1.86)
Vifit	0.03	0.00 - 0.06	-1.75	<0.001	(-1.86) - (-1.64)	0.48	0.42 - 0.53	-1.85	<.001	(-1.92) - (-1.79)
Pedometer	-0.31	(-0.49) - (-0.12)	-1.98	<0.001	(-2.01) - (-1.96)	0.24	0.14 - 0.34	-1.98	<.001	(-2.04) - (-1.95)
Accupedo	-0.02	(-0.09) - 0.05	-1.93	<0.001	(-1.99) - (-1.87)	-0.08	(-0.19) - 0.02	-1.98	<.001	(-2.02) - (-1.93)
Pedometer 2.0	-0.40	(-0.56) - (-0.24)	-1.99	<0.001	(-2.02) - (-1.96)	-0.23	(-0.37) - (-0.10)	-2.01	<.001	(-2.04) - (-1.97)

Note. CI: confidence interval; diff: difference

Discussion

The aim of the present study was to examine the accuracy of a variety of PA monitors and smartphone apps in measuring steps and distance during self-paced outdoor walking and running in a sample of healthy adolescents. To our knowledge, this is the first study to examine these estimates from competing technologies, including both GPS-accelerometer monitors and smartphone apps, in youth. The primary finding regarding step count was that accelerometer-based monitors (Vivofit and Vifit) were more accurate than smartphone apps in both walking and running conditions, and only the Accupedo app had an acceptable MAPE (≈10%) to be considered valid for use by the general population. Regarding distance validation, in which GPS and accelerometer-based monitors and apps were tested, it was found that only the GPS monitor and apps were valid in both conditions, with low individual errors (<10%). Accelerometer-based monitors and apps had high MAPE, except for Vifit and Accupedo during walking. An important finding was that all GPS freeware apps had comparable accuracy levels with the PA monitor in distance travelled, however only one app (i.e., Accupedo) had comparable accuracy to monitors for step count.

Due to a lack of studies in adolescents, comparisons for the PA monitors and apps from the current study were limited to studies in adult populations. Most previous reviews of validation studies for step count concluded that PA monitors were likely to meet acceptable accuracy levels (Dowd et al., 2018; Feehan et al., 2018; Evenson & Spade, 2020; Moore et al., 2020). Also, hip-worn PA monitors had greater accuracy in measuring steps than wrist-worn activity trackers (Gaz et al., 2018; Moore et al., 2020), and accuracy increased with increasing walking speed (Huang, Xu, Yu, & Shull, 2016; Höchsmann et al., 2018).

The results of the current study partially supported previous findings. Vivofit, which is a wrist-worn PA monitor, was more accurate in both conditions than Vifit (hip-based tracker), and Vivofit had increased validity during running, compared to walking. On the other hand, Vifit performed accurately during walking, with minimum group- and individual-level errors, however validity decreased when speed increased. A possible explanation for these inconsistent findings with previous studies, as mentioned by Adamakis (2020), is that Vifit uses one step-detection algorithm and this algorithm does not differentiate between various activities (i.e., walking vs running), failing to take into account the increased step length during running. Vifit can be considered suitable only for light activities in adolescents, such as brisk walking, while Vivofit is more accurate. Considering previous studies regarding Vivofit's step count validity (e.g., Huang et al., 2016; Höchsmann et al., 2018), this monitor can be used as an outcome measure within a clinical trial, while Vifit is valid for use by the general population.

A unique aspect of this validation study was the inclusion of both monitors and freeware GPS- and accelerometer-based apps. In general, GPS monitor and apps outperformed accelerometer-based monitors and apps for distance validity. Previous studies have shown that GPS devices provide accurate estimates of distance travelled during PA (Scott, Scott, & Kelly, 2016;

Adamakis, 2017; Tierney & Clarke, 2019), and Gray, Jenkins, Andrews, Taaffe, & Glover (2010) noted that GPS were more accurate in tracking distances during linear activities compared to nonlinear activities. This was supported by current results, as all GPS devices had low group- and individual-level errors. More specifically, Forerunner and Sports Tracker had comparable MAPE <5%, while Runkeeper and Runtastic had MAPE ≤10%, for both walking and running conditions.

On the contrary, accelerometer-based monitors and apps had higher distance errors, and only Vivofit had MAPE <15% in both conditions. Vifit and Accupedo had low MAPE during walking (≤10%), however their distance validity decreased significantly during running. The two remaining apps (Pedometer and Pedometer 2.0) had large individual and group errors for both walking and running conditions (>20%). This finding is consistent with previous studies, as most activity trackers and apps have been found accurate with step count but lacked accuracy in reporting distance (Evenson, Goto, & Furberg, 2015; Evenson & Spade, 2020; Gaz et al., 2018; Silva et al., 2020). This error in distance estimation might be a result of inaccurate initial step detection, inappropriate transformation algorithm(s) from step count into distance, and/or step length variability during PA. For example, Vivofit, which was the most accurate PA monitor for step count, was also the most accurate monitor for distance travelled, even though the error for distance estimation was significantly increased.

Regarding step count from accelerometer-based apps, it was previously concluded that there is conflicting and insufficient evidence on the validity and reliability of various apps for measuring PA, and speed and the place where the smartphone was carried had an impact on validity (Funk et al., 2019; Silva et al., 2020). Some studies showed that accelerometer apps slightly differed in their accuracy of step detection and generally demonstrated good validity (Case, Burwick, Volpp, & Patel, 2015; Höchsmann et al., 2018), while in other studies the apps showed moderate validity and did not meet the recognized standards (Orr et al., 2015). In the present study, Accupedo was the most accurate app for step count with acceptable errors (≈10%), while Pedometer and Pedometer 2.0 apps had high errors (>30%) in both conditions. Accupedo was also the only app with comparable validity to the monitors' validity, suitable for use by the general population. These differences are mainly because monitors and apps use proprietary methods to detect steps and hence, differences may exist in the types of movements that are captured as steps, and variability in step output is caused (John, Morton, Arguello, Lyden, & Bassett, 2018). The smartphones' position did not impact step detection accuracy, as all smartphones were placed close to the body, around the waist. We are uncertain whether the accuracy would improve if smartphones were placed in a different position, i.e., around the arm.

The main strengths of this study included the selection of monitors using various technologies (i.e., accelerometry and GPS), and the comparison to a criterion measure. Other strengths included a sample consisting of adolescents, even distribution of boys and girls, submaximal outdoor walking and running tests in a realistic setting and randomization of the two activities to prevent systematic bias in the measurement. Also, the running activity was performed at a high speed. Limitations to this study included the sample size consisting of healthy participants, while future studies should include more semi- or un-structured activities in free-living environment. In addition, future studies should examine the validity of apps during activities of daily living, preferably over a time frame of 2-4 days to assess the suitability of these devices to be used for long-term accelerometry. Finally, the role of smartphone's optimal position on the human body during exercise should be further investigated.

Conclusion

PA tracking devices have been shown to increase daily PA, but the reliability and validity of numerous commercially available monitors and apps remain unclear. In this validation study, each device returned some level of consistency and accuracy during outdoor walking and running in adolescents, GPS monitor and apps were deemed to be valid for distance during for both conditions tested. Forerunner and Sports Tracker showed comparable accuracy levels, suitable for use as an outcome measure within a clinical trial, while the two remaining GPS apps can be used by the general population. Accelerometer-based monitors and apps were not suitable for measuring distance; however, these monitors were valid for step count and only Accupedo app provided similar estimates. Similar studies should be continuously conducted as new fitness trackers, smartwatches and apps are released to the consumer market every year.

Acknowledgements
There are no acknowledgments.

Conflict of Interest
The author declares that there is no conflict of interest.

Received: 15 March 2022 | **Accepted:** 5 May 2022 | **Published:** 01 June 2022

References

Adamakis, M. (2017). Comparing the validity of a GPS monitor and a smartphone application to measure physical activity. *Journal of Mobile Technology in Medicine, 6*(2), 28-38. https://doi.org/10.7309/jmtm.6.2.4

Adamakis, M. (2020). Criterion validity of wearable monitors and smartphone applications to measure physical activity energy expenditure in adolescents. *Sport Sciences for Health, 16*(4), 755-763. https://doi.org/10.1007/s11332-020-00654-2

Adamakis, M. (2021). Criterion validity of iOS and Android applications to measure steps and distance in adults. *Technologies, 9*(3), 55. http://dx.doi.org/10.3390/technologies9030055

American College of Sports Medicine (2006). *ACSM's guidelines for exercise testing and prescription* (7th ed.) Baltimore, MD: Lippincott Williams and Wilkins.

Bent, B., & Dunn, J. P. (2020). Wearables in the SARS-CoV-2 pandemic: what are they good for? *JMIR mHealth and uHealth, 8*(12), e25137. https://doi.org/10.2196/25137

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 327*(8476), 307-310. PMID: 2868172

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., … Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British Journal of Sports Medicine, 54*(24), 1451-1462. https://doi.org/10.1136/bjsports-2020-102955

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA, 313*(6), 625-626. https://doi.org/10.1001/jama.2014.17841

Dowd, K. P., Szeklicki, R., Minetto, M. A., Murphy, M. H., Polito, A., Ghigo, E., … & Donnelly, A. E. (2018). A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. *The International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 15. https://doi.org/10.1186/s12966-017-0636-2

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 159. https://doi.org/10.1186/s12966-015-0314-1

Evenson, K. R., & Spade, C. L. (2020). Review of validity and reliability of Garmin activity trackers. *Journal for the Measurement of Physical Behaviour, 3*(2), 170-185 https://doi.org/10.1123/jmpb.2019-0035

Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., Hamilton, C. B., & Li, L. C. (2018). Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth, 6*(8), e10527. https://doi.org/10.2196/10527

Funk, M. D., Salazar, C. L., Martinez, M., Gonzalez, J., Leyva, P., Bassett Jr., D., & Karabulut, M. (2019). Validity of smartphone applications at measuring steps: does wear location matter? *Journal for the Measurement of Physical Behaviour, 2*(1), 22-27. https://doi.org/10.1123/jmpb.2018-0025

Gaz, D. V., Rieck, T. M., Peterson, N. W., Ferguson, J. A., Schroeder, D. R., Dunfee, H. A., Henderzahs-Mason, J. M., & Hagen, P. T. (2018). Determining the validity and accuracy of multiple activity-tracking devices in controlled and free-walking conditions. *American Journal of Health Promotion, 32*(8), 1671-1678. https://doi.org/10.1177/0890117118763273

Gray, A. J., Jenkins, D., Andrews, M. H., Taaffe, D. R., & Glover, M. L. (2010). Validity and reliability of GPS for measuring distance travelled in field-based team sports. *Journal of Sports Sciences, 28*(12), 1319-1325. https://doi.org/10.1080/02640414.2010.504783

Höchsmann, C., Knaier, R., Eymann, J., Hintermann, J., Infanger, D., & Schmidt-Trucksäss, 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. https://doi.org/10.1111/sms.13074

Huang, Y., Xu, J., Yu, B., & Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait and Posture, 48*, 36-41. https://doi.org/10.1016/j.gaitpost.2016.04.025

John, D., Morton, A., Arguello, D., Lyden, K., & Bassett, D. (2018). "What Is a step?" Differences in how a step is detected among three popular activity monitors that have impacted physical activity research. *Sensors, 18*(4), 1206. https://doi.org/10.3390/s18041206

Johnston, W., Judice, P. B., Molina García, P., Mühlen, J. M., Lykke Skovgaard, E., Stang, J., … & Sardinha, L. B. (2020). Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *British Journal of Sports Medicine, 55*(14), 780-793. https://doi.org/10.1136/bjsports-2020-103147

Keadle, S. K., Lyden, K. A., Strath, S. J., Staudenmayer, J. W., & Freedson, P. S. (2019). A framework to evaluate devices that assess physical behavior. *Exercise and Sport Sciences Reviews, 47*(4), 206-214. https://doi.org/10.1249/JES.0000000000000206

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clinical and Experimental Pharmacology and Physiology, 29*(7), 527-536. https://doi.org/10.1046/j.1440-1681.2002.03686.x

Moore, C. C., McCullough, A. K., Aguiar, E. J., Ducharme, S. W., & Tudor-Locke, C. (2020). Toward harmonized treadmill-based validation of step-counting wearable technologies: a scoping review. *Journal of Physical Activity and Health, 17*(8), 840-852. https://doi.org/10.1123/jpah.2019-0205

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise, 48*(8), 1619-1628. https://doi.org/10.1249/MSS.0000000000000933

Omura, J. D., Carlson, S. A., Paul, P., Watson, K. B., & Fulton, J. E. (2017). National physical activity surveillance: users of wearable activity monitors as a potential data source. *Preventive Medicine Reports, 5*, 124-126. https://doi.org/10.1016/j.pmedr.2016.10.014

Orr, K., Howe, H. S., Omran, J., Smith, K. A., Palmateer, T. M., Ma, A. E., & Faulkner, G. (2015). Validity of smartphone pedometer applications. *BMC Research Notes, 8*(1), 733. https://doi.org/10.1186/s13104-015-1705-8

Pobiruchin, M., Suleder, J., Zowalla, R., & Wiesner, M., (2017). Accuracy and adoption of wearable technology used by active citizens: a marathon event field study. *JMIR mHealth uHealth, 5*(2), e24. https://doi.org/10.2196/mhealth.6395

Scott, M. T., Scott, T. J., & Kelly, V. G. (2016). The validity and reliability of Global Positioning Systems in team sport: a brief review. *Journal of Strength and Conditioning Research, 30*(5), 1470–1490. https://doi.org/10.1519/JSC.0000000000001221

Silva, A. G., Simões, P., Queirós, A., Rodrigues, M., & Rocha, N. P. (2020). Mobile apps to quantify aspects of physical activity: a systematic review on its reliability and validity. *Journal of Medical Systems, 44*(2), 51. https://doi.org/10.1007/s10916-019-1506-z

Tierney, P., & Clarke, N. (2019). A comparison of a smartphone app with other GPS tracking type devices employed in football. *Exercise Medicine, 3*(4), 1-7

. https://doi.org/10.26644/em.2019.004

Troiano, R. P., Stamatakis, E., & Bull, F. C. (2020). How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. *British Journal of Sports Medicine, 54*(24), 1468-1473. https://doi.org/10.1136/bjsports-2020-102621

Warner, R. M. (2012). *Applied Statistics: From Bivariate Through Multivariate Techniques* (2nd ed.). Los Angeles, CA: Sage.

WHO (2020). *Guidelines on physical activity and sedentary behaviour*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO. Retrieved online from https://www.who.int/publications/i/item/9789240015128