Smartphone apps for tracking physical activity and sedentary behavior: A criterion validity review

Aplicativos de smartphone para monitorar a atividade física e o comportamento sedentário: uma revisão sistemática de estudos de validação

AUTHOR'S

Raul Cosme Ramos Prado[1,2] (ID)
Margarethe Thaisi Garro Knebel[1,3] (ID)
Evelyn Helena Corgosinho Ribeiro[1] (ID)
Inaian Pignatti Teixeira[1] (ID)
Jeffer Eidi Sasaki[4] (ID)
Luciano Vieira de Araújo[5] (ID)
Paulo Henrique Guerra[1,6] (ID)
Alex Antonio Florindo[1,3,5] (ID)

1 University of Sao Paulo, Physical Activity Epidemiology Group, Sao Paulo, São Paulo, Brazil.

2 University of São Paulo, School of Physical Education and Sport, São Paulo, São Paulo, Brazil.

3 University of Sao Paulo, Graduate Program in Nutrition in Public Health, School of Public Health, Sao Paulo, São Paulo, Brazil.

4 Federal University of Triangulo Mineiro, Graduate Program in Physical Education, Uberaba, Minas Gerais, Brazil.

5 University of Sao Paulo, School of Arts, Sciences and Humanities, São Paulo, São Paulo, Brazil.

6 Federal University of Fronteira Sul, Chapecó, Santa Catarina, Brazil.

CORRESPONDING

Raul Cosme Ramos Prado
pradorcr@gmail.com
University of Sao Paulo, Brazil. Address: Rua Arlindo Béttio, 1000 - Vila Guaraciaba, São Paulo, São Paulo,
ZIP CODE: 03828-000.

DOI

10.12820/rbafs.27e0270

ABSTRACT

Smartphone apps have been developed and investigated in validation studies for tracking human behavior such as physical activity (PA) and sedentary behavior (SB). However, as it is unclear whether these apps are valid for tracking PA and SB when compared to research-grade accelerometers, thus, this systematic review aimed to investigate the validity of smartphone apps for tracking PA and SB using the accelerometer as a criterion measure. A systematic search was conducted in PubMed, Web of Science, SportDiscus, and Scopus databases. The mean percentage difference (MPD) was used to evaluate criterion validity. Ten studies (n = 662) validating different apps using ActiGraph accelerometers as criteria measure (six were conducted in free-living conditions, two in laboratory conditions, and two in both conditions) were included for analyses. While four apps were considered valid for tracking PA, six were not valid or fully valid. The MPD analysis revealed that apps provide no valid scores for tracking PA measures (MPD = -12.6 – 37.7). The scarcity of studies investigating SB limits the tracking of the results on this behavior. Study designs, smartphone location, and exercise intensity tend to affect the accuracy of apps tracking PA; thus, the current review showed conflicting results among studies. This review shows that it is not possible to generalize the valid scores for all apps.

Keywords: Accelerometry; Measurement equipment; Sitting position; Epidemiology.

RESUMO

Aplicativos para smartphones têm sido desenvolvidos e investigados em estudos de validação para rastreamento de comportamento humano, como atividade física (AF) e comportamento sedentário (CS). No entanto, como não está claro se esses aplicativos são válidos para rastrear AF e CS quando comparados a acelerômetros de grau de pesquisa, portanto, essa revisão sistemática teve o objetivo investigar a validade de aplicativos de smartphone para rastreamento de AF e CS usando o acelerômetro como medida de critério. Uma busca sistemática foi realizada em quatro bases de dados. A diferença percentual média (MPD) foi utilizada para avaliar a validade de critério. Dez estudos (n = 662) validando diferentes aplicativos usando acelerômetros ActiGraph como medida de critério (seis foram realizados em condições de vida diária, dois em condições de laboratório e dois em ambas as condições) foram incluídos para análise. Enquanto quatro aplicativos foram considerados válidos para rastreamento de AF, seis não eram válidos ou totalmente válidos. A análise do MPD revelou que os aplicativos não fornecem pontuações válidas para rastrear medidas de AF (MPD = -12,6 – 37,7). A escassez de estudos investigando o CS limita o rastreamento dos resultados sobre esse comportamento. Desenhos de estudo, localização do smartphone e intensidade do exercício tendem a afetar a precisão dos aplicativos que rastreiam AF; assim, a presente revisão mostrou resultados conflitantes entre os estudos. Esta revisão mostra que não é possível generalizar as pontuações válidas para todos os aplicativos.

Palavras-chave: *Acelerometria; Equipamentos de medição; Postura sentada; Epidemiologia.*

Introduction

Quantifying physical activity (PA) and sedentary behavior (SB) allows scientists to understand many dose-response relationships with health outcomes such as chronic diseases[1] and mortality[2]. This knowledge is the background for developing global guidelines, including the World Health Organization recommendations[3]. In this sense, measuring PA and SB has been a matter of accuracy for providing high-quality evidence, and several methods of measurement have been proposed and revised over the years[4,5]. Self-report questionnaires are an inexpensive data collection method and usually demons-

trate good agreement to track moderate- and vigorous--intensity PA[6]; however, this method proves inadequate for tracking light-intensity PA and sedentary activities[7–9].

Conversely, accelerometers are considered an accurate method for tracking PA at different intensities (e.g., light-, moderate-, and vigorous-intensity)[10]. Using concepts adopted by early accelerometers designed to assess vibrations and detect motion in industrial settings[11], these devices have been adapted and commercialized worldwide[10,12] for tracking human movement. Owing to technological developments, accelerometers have become sufficiently small and lightweight, manufactured in the shape of wearable devices capable of measuring PA in free-living settings. However, research-grade accelerometers are relatively expensive, and some require the manufacturers' software to process data. Thus, despite several studies using accelerometers, the high costs associated with using such devices may often hamper their use in epidemiology research, as these studies typically comprise large sample sizes[13].

Recently, the fitness market has promoted the proliferation of commercial wearable devices for tracking human movement by combining inertial sensors, gyroscopes, and GPS. For example, Fitbit devices are among the most widely used wearable activity trackers worldwide[14]. Nevertheless, a recent review showed a tendency of Fitbit devices to underestimate step count in controlled trials and overestimate step count in free-living trials. Moreover, the authors also highlighted that the accuracy of Fitbit devices to track distance and energy expenditure is poor[15].

Another alternative to wearable activity trackers is smartphone devices. Currently, several apps are developed for smartphones to evaluate the PA and SB of users. Approximately 2.5 billion people possessed a smartphone in 2016. Since then, this figure has increased by 40%, meaning that 44% (3.8 billion) of people worldwide currently possess a smartphone[16]. Statistics also indicate that 7.33 billion people worldwide will own a mobile device in 2023[16]. This scenario will facilitate the development of epidemiological studies to track PA and SB in large samples by using smartphone apps. However, smartphone apps face challenges for PA and SB assessment in terms of scientific research verifying their validity and reliability.

For example, a recent systematic review by Silva et al.[17] synthesized findings of studies that tested the validity and reliability of smartphone apps for monitoring PA. The authors synthesized results related to

several criterion measures by smartphone apps (i.e., pedometer, manual step counting by the researchers and self step count by the participants, gait analysis, video observation, respiratory gas analyzer, GPS, pre-established distance, and travel on a treadmill). Conflicting results among studies were observed, which ranged from poor to excellent scores for both validity and reliability levels of smartphone apps for monitoring PA and SB, leading the authors to suggest that low methodological quality may be the cause of this ambiguity. However, this review does not clarify[17] the accuracy of apps for tracking PA as compared to that of a research-grade accelerometer for measuring light-, moderate-, and vigorous-intensity physical activities. Additionally, the authors did not focus on studies that validated smartphone apps to monitor SB.

Therefore, the present current systematic review aimed to summarize evidence of the validity of smartphone apps for tracking PA and SB using the accelerometer as a criterion measure.

Methods

This systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) platform (CRD42020193727) and was conducted according to the guidelines by Cochrane[18] and Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA)[19]. The PICOS (Population [humans without specific restrictions], Intervention [smartphone apps], Comparison [research-grade accelerometers], Outcome [PA and SB measurements], and Study [laboratory or free-living]) criteria[20] were used for assessing the eligibility of studies included in this review and for conducting the methodological assessment following the Cochrane Handbook for Systematic Reviews of Interventions guidelines[21].

The present systematic review included only peer-reviewed original studies aimed at testing the criterion validity of the PA and SB measurements derived from smartphone apps against that of research-grade accelerometers. There were no restrictions concerning the publication date of the studies or the age range and the physical or clinical status of the samples. Studies were excluded according to the following criteria: (i) the reference standard was not an accelerometer; (ii) no PA or SB measures; (iii) studies focused on smartphone apps requiring external components or sensors (e.g., watchbands or bracelets); (iv) no smartphone app being validated; (v) the samples were not human; (vi)

studies written in languages other than English; and (vii) conferences, reviews, or case studies.

The screening of studies was independently carried out by two researchers (RP and MK) using Mendeley Desktop software (v1.19.4, 2008–2019), and the opinion of the senior researcher (AAF) of the study was considered in case of disagreements.

The searches were conducted on February 19, 2021 in the electronic databases of PubMed, Web of Science, SportDiscus, and Scopus. Additionally, reference lists of the included articles were examined. In each database, two strategies were implemented, based on the following strategy elaborated for PubMed:

i) PA: ((((physical activity* OR exercise OR fitness OR Sport*) AND (mobile app OR wearable devices OR mobile phones OR mobile phone application OR smartphone OR smartphone application OR mobile technology OR smartphone technology)) AND (accelerometer OR accelerometry OR acceleration OR counts OR ActiGraph OR RT3 OR inclinometer OR Activpal OR pedometer OR doubly labeled water OR heart rate))) AND (validation OR validate* OR measurement OR software validation).

ii) SB: ((((sedentar* OR sedentary behavior OR physical inactivit*)) AND (mobile app OR wearable devices OR mobile phones OR mobile phone application OR smartphone OR smartphone application OR mobile technology OR smartphone technology)) AND (accelerometer OR accelerometry OR acceleration OR counts OR ActiGraph OR RT3 OR inclinometer OR Activpal OR pedometer OR doubly labeled water OR heart rate))) AND (validation OR validat* OR measurement OR software validation).

Independently, two researchers (RP and MK) extracted and synthesized the following information from the selected articles: i) characteristics of subjects (i.e., sample size, age range); ii) physical profile (e.g., physically activity, sedentary); iii) smartphone and accelerometer information; iv) validation design between the smartphone app and accelerometer; v) outcomes accuracy (validity) and consistency (reliability) with respect to PA and SB measures (e.g., step counts, metabolic equivalents (METs) sit-to-stand, acceleration, counts). Subsequently, data extraction results were compared to verify inconsistency conflicts.

RP and MK assessed the methodological aspects of the included studies based on the critical appraisal tool for validity and reliability of objective clinical tools developed by Brink and Louw[22]. This tool classifies the

methodological quality, scoring 13 items as "yes," "no" or "not applicable." Adopting similar methods, in Silva et al.[17] methodological quality was calculated and presented according to the percentage of "yes" responses for applicable items. The percentage of agreement between researchers was calculated, and the researchers conducting this review held a virtual meeting to discuss disagreements.

Interrater reliability analyses for the methodological assessment were conducted using the kappa statistic. The statistical criteria for assessing the validity of the apps were: i) mean percent difference (MPD, <10%), ii) kappa coefficient (>0.20), iii) correlation coefficient (ICC), and iv) simple correlation (>0.70)[23,24]. Unreported MPD values were calculated for studies that presented the average accelerometer and smartphone app outcomes. Similarly, this strategy was conducted in a previous review[17] using a standardized equation[25]:

$$\left(\frac{\text{smartphone app measure} - \text{accelerometer measure}}{\text{accelerometer measure}} \right) \times 100$$

The GetData Graph Digitizer (v 2.26.0.20) was used to extract information in the studies that reported their data only in figures to calculate MPD.

Results

Search results and study characteristics

The current systematic search identified 2814 studies in selected databases. After excluding 771 duplicates, 2043 studies remained and were subjected to title and abstract screening. According to the exclusion criteria, 1793 studies were removed during the screening of title and abstract, and 240 studies were further excluded after reading the full text. Thus, the remaining 10 studies[26–35] were included in this review for conducting descriptive synthesis (Supplementary Data 1).

The selected studies were published between 2013 and 2020 and were conducted in various countries, most originating from the US[27,32,33], followed by Spain[28,34], as well as two multicenter studies[29,33]. A total of 662 (50% males) individuals participated in these studies, and the sample size ranged from 21[26] to 158 participants[29]. All studies included samples for convenience, and the participants were mostly healthy (seven out of 10 studies), except for three studies that included people with noncommunicable diseases[27,30,35]. The participants were exclusively adults and the mean age ranged from 25.9±9.4 years[31] to 69.5±13.1 years[27] (Table 1). Five studies[28,31,33–35] were composed of par-

ticipants with normal weight according to body mass index (BMI); others[27,29,30,32] included overweight participants (BMI > 25 kg/m²), and one study[26] did not describe anthropometric data of the sample.

Most of the studies used commercial smartphone Android apps[28,29,32–35]. The following tested for criterion validity: Moves[26], VascTrac[27], CalFit[28], ExpoApp[29], Walkmeter[30], Health app[31], and Movn[33]. The URLs for downloading the smartphone apps are provided by the corresponding authors and are shown in Supplementary Data 1. The smartphone was carried at the waist-hip level (e.g., in a pocket or belt), except for two studies in which the participants held a smartphone in their hand in laboratory conditions[27,31]. The processing and conversion of acceleration data were performed by using different techniques. Two studies did not report such information[30,31] (Table 2).

Six apps[28,29,32–35] were able to measure the intensity of PA. The CalFit app provided PA intensity through acceleration or MET outputs, and distance measures. Only the customized app (smartphone app developed by researchers) from Hekler et al.[32] quantified both SB and PA measures (Table 2).

ActiGraph accelerometers were used in all included studies, in which the GT3X[26,28] and GT3X+[31,32,34,35] models appeared in three studies each. The accelerometers were exclusively attached on the hip/waist, mainly on the right side of the body. The sampling frequency of ActiGraph accelerometers ranged between 30 to 100 Hz for six studies[27,28,31,32,34] (Table 2).

Study designs and result synthesis

Six studies were conducted in free-living conditions[26,28–31,35], two studies were conducted in laboratory conditions designed by researchers[27,34], and the remaining two studies were conducted in both conditions[32,33]. Most studies adopted the cutoff values by Freedson et al.[36] for PA classification (light, moderate, and vigorous physical activities). The cutoff values by Matthews et al.[37] (counts<100/min) were used to characterize SB. It was observed that the design adopted by studies belonged to either of the following two categories: studies with only validity design[26–29,33–35] and studies with validity and reliability design[30–32] (Figure 1).

Although step count measure[26,27,30,31,35] was the most frequent measurement compared between the smartphone apps and accelerometer in the studies, validation analyses were also conducted for METs[28,29,33], duration of PA[28], raw acceleration, and counts per minute[28,32,34] (Table 3).

Methodological assessment

Good agreement (k = 83%) between raters was observed for the methodological assessment. The studies showed good quality assessment (60%), followed by low-quality assessment (13%), and assessment does not apply (27%). Studies showed good quality assessment for items #1, #3, #7, and #9 to #13. Low quality (< 50%) was observed for #2 (researcher experience). The data are reported in Figure 1.

Step Counts

Two studies conducted in free-living conditions[26,31] showed that smartphone apps underestimated the

Table 1 – Characteristics of studies.

Reference	Country	Sample	Mean age ± SD
Asimina et al.[26]	GRC	21 participants	No information
Ata et al.[27]	USA	114 (88 ♂ and 26 ♀) peripheral artery disease patients	69.5 ± 13.1
Donaire-Gonzalez et al.[28]	SPA	36 (13 ♂ and 23 ♀) healthy participants	31.0 ± 8.0
Donaire-Gonzalez et al.[29]	NL, SWI, UK, and ITA	158 (91 ♂ and 97 ♀) adults	60.5 ± 6.5
Douma et al.[30]	NL	64 (40 ♂ and 24 ♀) patients with cancer	63.0 ± 11.5
Duncan et al.[31]	CA	33 (11 ♂ and 22 ♀) healthy young adults	25.9 ± 9.4
Hekler et al.[32]	USA	38 healthy participants = 15 participants (7 ♂ and 8 ♀) in laboratory study and 23 participants (6 ♂ and 17 ♀) in free-living study	laboratory (55.5 ± 6.6) and free-living (57.0 ± 6.4) study
Maddison et al.[33]	IE and USA	29 adults = 21 participants (21 ♂ and 13 ♀) in laboratory study and 8 participants in free-living study	laboratory study (27.0 ± 7.9)
Rodriguez et al.[34]	SPA	32 (19 ♂ and 13 ♀) participants	27.8 ± 12.9
Zhai et al.[35]	DE	137 adults = 70 healthy participants (23 ♂ and 47 ♀) and 67 participants with multiple sclerosis (25 ♂ and 42 ♀)	healthy participants (41.5 ± 12.8) and participants with multiple sclerosis (42.9 ± 10.9)

♂ = men; ♀ = women; SD = standard deviation; BMI = body mass index; DE = Germany; GRC = Greece; USA = United States of America; SPA = Spain; NL = Netherlands; SWI = Switzerland; UK = United Kingdom; ITA = Italy; CA = Canada; IE = Ireland.

Table 2 – Smartphone, apps, and criterion measure information.

	Smartphone		App			Accelerometer		
Reference	Model	Side and location attached	App name/ mobile operating system	Sensors/sampling frequency	Enabled measures in-app	Model	Side and location attached	Sampling frequency
Asimina et al.[26]	NI	pocket or bag	Moves/NI	NI	step counts and distance	ActiGraph GT3X	right hip	NI
Ata et al.[27]	iPhones SE, 6, 7, and 7 Plus	hand, shirt front, and pants pockets	VascTrac/iOS	NI	step counts and distance	ActiGraph GT9X	right hip	100 Hz
Donaire-Gonzalez et al.[28]	Google G1	the frontal mean points between both anterior superior iliac spines	CalFit/Android	tri-axial accelerometers/ 10 Hz	PA, METs, duration of PA and acceleration	ActiGraph GT3X	right hip	30 Hz
Donaire-Gonzalez et al.[29]	Samsung Galaxy S3	waist	Expo/Android	accelerometer/ 10 s	PA and METs	wGT3X+ (in ITA, UK and SWI) and wActiSleep+ (in NL)	waist	NI
Douma et al.[30]	iPhone SE	hip-waist or attached to a belt	Walkmeter/iOS	accelerometer/ NI	step counts and distance	ActiGraph wGT3X	waist	NI
Duncan et al.[31]	iPhone 6 (n = 13), 6S (n = 8), 6S+ (n= 1), SE (n = 8), 7 (n = 2) and 7+ (n = 1)	left hand	Health/iOS	accelerometer and pedometer	step counts and counts	ActiGraph GT3X+	right hip	30 Hz
Hekler et al.[32]	HTC MyTouch, Google Nexus One, and Motorola Cliq	laboratory: non-dominant hip; free-living: hip or pockets	Custom app/ Android	accelerometer/20 Hz for the HTC MyTouch and Google Nexus One; 80 Hz for the Motorola Cliq	SB and PA	ActiGraph GT3X+	non-dominant hip	80 Hz
Maddison et al.[33]	Phase 1: Moto G; phase 2: Samsung Galaxy Nexus S	phase 1: right hip; phase 2: right iliac crest	Movn/Android and iOS	accelerometer/ 50 Hz	PA and METs	ActiGraph GT1M	right hip	30 Hz
Rodriguez et al.[34]	Samsung Galaxy Trend Plus GT-S7580	right pocket and right hip	Custom app/ Android	accelerometer/ 30 Hz	PA and counts	ActiGraph GT3X+	right hip	30 Hz
Zhai et al.[35]	Samsung Galaxy S4 mini	habitual position	Custom app/ Android	Accelerometer/ 2 Hz	Acceleration	GT3X+	non-dominant wrist	NI

NI = No information; PA = physical activity; SB = sedentary behavior; METs = metabolic equivalent; NL = Netherlands; SWI = Switzerland; UK = United Kingdom; ITA = Italy.

daily step count as compared to that of the ActiGraph accelerometer (Moves: MPD = -32.9, and Health: MPD = -21.5, bias = -1341.8, LoA = -4297.1–1613.6) and no acceptable correlation coefficient (Custom app: r = -0.12–0.47) was observed in one study[35]. Two smartphone apps were considered valid for step count measure (MPD <10% and ICC > 0.70), of which one was examined in free-living conditions (Walkmeter)[30] and the other was tested in laboratory conditions (Vasc-

Trac)[27]. A valid study conducted in free-living conditions[30] also conducted a test-retest reliability analysis, which showed acceptable reliability (MPD = -8.6, ICC = 0.91, bias = 43.14, LoA = -1273.0–1373.0) (Table 3).

METs

Based on the free-living design, a study[33] showed that the smartphone app Movn was not valid for measuring METs during light- to very-vigorous PA (MPD =

A. Summary of critical appraisal

Study	Valid and/or Reliable	Critical Appraisal
		1 2 3 4 5 6 7 8 9 10 11 12 13
Asimina et al.[26]	Valid	− − + ? ? ? + ? + + + + −
Ata et al.[27]	Valid	+ − + ? ? ? + ? + + + + +
Donaire-Gonzalez et al.[28]	Valid	+ − + ? ? ? + ? + + + + +
Donaire-Gonzalez et al.[29]	Valid	+ − + ? ? ? + ? + + + − +
Douma et al.[30]	Valid/reliable	+ − + + ? ? + + + + − + +
Duncan et al.[31]	Valid/reliable	+ − + ? ? ? + − + + + + +
Hekler et al.[32]	Valid/reliable	+ − + ? ? ? + − + + + + +
Maddison et al.[33]	Valid	+ − + ? ? ? + ? + + + + +
Rodriguez et al.[34]	Valid	+ − + ? ? ? + ? + + + + +
Zhai et al.[35]	Valid	+ − + ? ? ? + ? + + + + −

B. Critical appraisal graph
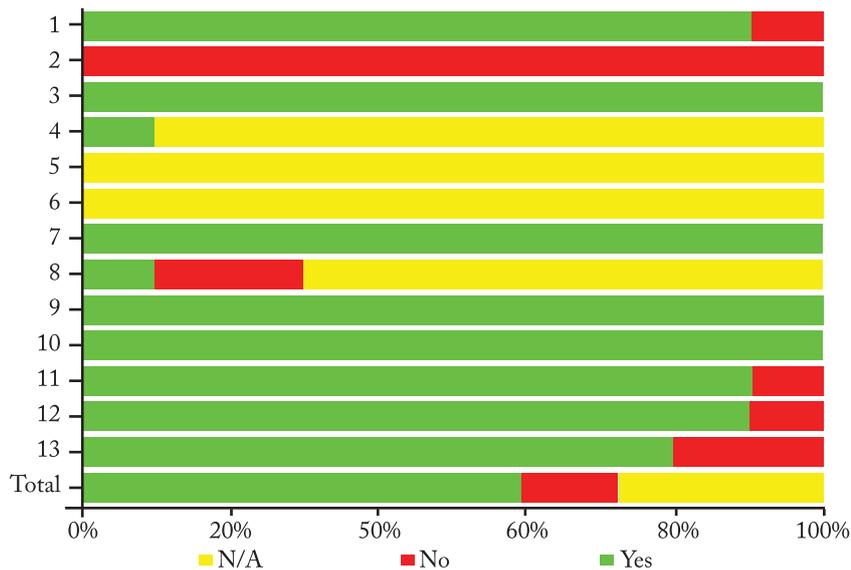
Figure 1 – Summary (panel A) and graph (panel B) of critical appraisal of selected articles for reliability and validity. Green = Yes, Red = No, Yellow = not applicable (N/A). (#1) If human subjects were used, did the authors give a detailed description of the sample of subjects used to perform the (index) test on?; (#2) Did the authors clarify the qualification or competence of the rater(s) who performed the (index) test?; (#3) Was the reference standard explained?; (#4) If interrater reliability was tested, were raters blinded to the findings of other raters?; (#5) If intrarater reliability was tested, were raters blinded to their prior findings of the test under evaluation?; (#6) Was the order of examination varied?; (#7) If human subjects were used, was the period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?; (#8) Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval between repeated measures?; (#9) Was the reference standard independent of the index test?; (#10) Was the execution of the reference standard described in sufficient detail to permit its replication?; (#11) Was the execution of the reference standard described in sufficient detail to permit its replication?; (#12) Were withdrawals from the study explained?; (#13) Were the statistical methods appropriate for the study?

50.9–129.7). Another study[28] found similar results for vigorous-intensity PA (MPD = 17.7) with the smartphone attached to the waist; however, this study was valid only for light- and moderate-intensity PA (MPD = -1.4–3.2). A multicenter study[29] showed that the smartphone app Expo was valid for assessing METs (concordance correlation coefficient = 75.7% and 89.9% for specific city algorithms); however, it did not test specific PA intensity. Based on a laboratory design, a study[33] assessed PA in a controlled treadmill speed. In this study, the overall analysis revealed that the smartphone attached at the hip level overestimated METs (MPD = 13.9) (Table 3).

Other measures

Two custom apps were considered valid for assessing counts per minute in laboratory conditions[32,34] attaching a smartphone to the right pocket and right hip (r = 0.77–0.94, bias = 1.99, LoA = -5.8–5.7). A study with a free-living design[32] was valid for counts per minute during SB (bias = -26.0, LoA = -279.5–227.6) and moderate-to-vigorous physical activity (MVPA) (bias = -1.3, LoA = -38.4–35.8). Acceptable validity (MPD < 10.0; r > 0.70) was observed for the duration of PA[28] (Table 3).

Results according to activity intensity

A study conducted in free-living design[32] presented an acceptable mean difference (bias = -26.0, LoA =

-279.5–227.6) for tracking SB based on the cutoff values by Matthews et al.[37] and MVPA (bias = -1.3, LoA = -38.4–35.8) based on the cutoff values by Freedson et al.[36], but none were valid for tracking light-intensity PA. The smartphone app CalFit[28] was valid for tracking light-intensity PA (MPD = 3.2) and moderate-intensity PA (MPD = -1.4) with PA based on the classification by Freedson et al.[36]. Measures for light-, moderate-, vigorous-, and very vigorous-intensity PA were not valid (MPD = 50.9–129.7) in a free-living study[33] based on the cutoff values by Freedson et al.[36]. However, based on Sasaki's cutoff values[38], a laboratory study[34] provided valid scores (r = 0.94; ICC = 0.94; bias = 1.99, LoA = -5.8–5.7; kappa = 0.87) for tracking light-, moderate-, vigorous-, and very vigorous-

Table 3 – Design and results of the studies.

| Reference	Setup	Outcome assessed	Cutoff	Statistical analysis	Primary results	Valid / Not valid
Studies conducted in free-living conditions (n = 6)						
Asimina et al.[26]	Free-living activity for a week.	Step counts	NI	MPD	Step counts showed no acceptable mean difference (MPD = -32.9) in average of days.	Not valid
Donaire-Gonzalez et al.[28]	Five consecutive days. Three days with at least 10 h of use was considered valid data.	Vertical axis counts, duration and METs	Freedson et al.[36]	MPD, Spearman's correlation coefficient and concordance correlation coefficient	METs showed acceptable mean difference at light-intensity (MPD = 3.2), moderate-intensity (MPD = -1.4), and no acceptable mean difference at vigorous-intensity (MPD = 17.7). Acceptable mean difference of active time above >1.5 METs (MPD = 2.2). Vertical axis counts showed acceptable correlation coefficient (r = 0.93). PA duration and METs showed acceptable concordance correlation coefficient (0.83–0.91%)	Light-intensity and moderate-intensity (valid) and vigorous-intensity (not valid)
Donaire-Gonzalez et al.[29]	Participants' location and PA were monitored three times for 24 h in three different seasons over one year. A minimum of 10 h of wearing time of devices was considered valid data.	METs	Crouter et al.[39]	Concordance correlation coefficient	METs showed acceptable concordance correlation coefficient (75.7%). METs showed acceptable concordance correlation coefficient in specific city algorithms (89.9%)	Valid
Douma et al.[30]	14 consecutive days during all waking hours, concomitant with a waist-worn accelerometer during the first seven days. A minimum of 10 h and four valid days per week of wearing time of devices were considered valid data.	Step counts	NI	MPD, ICC and Bland-Altman	Step counts showed acceptable mean difference (MPD = -0.6), correlation coefficient (ICC = 0.97 p < 0.05) and no systematic difference (bias = 43.14, LoA = -1273.0–1373.0). Step counts showed acceptable mean difference (MPD = -8.6) and correlation coefficient (ICC = 0.91 p < 0.05) in test-retest reliability analysis	Valid
Duncan et al.[31]	Three days of free-living protocol with at least 10 h of use was considered valid data.	Step counts	NI	MPD, ICC, Bland-Altman and t-tests	Step counts showed no acceptable error in the best day and average day (MPD = -18.2 and -21.5). Acceptable correlation coefficient in the best day and average day (ICC = 0.94 and 0.73, p < 0.05). Systematic differences in best day (bias = -982.4, LoA = -3242.4–1277.5), average day (-1341.8, LoA = -4297.1–1613.6) and statistical difference (p < 0.05)	Not valid

Continue…

Continue of **Table 3** – Design and results of the studies.

Reference	Setup	Outcome assessed	Cutoff	Statistical analysis	Primary results	Valid / Not valid
Zhai et al.[35]	Seven days of free-living.	Step counts/min and counts	NI	Spearman's correlation coefficient	Healthy participants: ActiGraph step counts/min showed no acceptable correlation with smartphone varVM (r = 0.47, p < 0.05) and with smartphone meanVM (r = 0.01, p > 0.05). Daily ActiGraph MVPA showed no acceptable correlation with smartphone varVM (r = 0.16, p > 0.05) and with smartphone meanVM (r = -0.16, p > 0.05). Participants with multiple sclerosis: ActiGraph step counts/min showed no acceptable correlation with smartphone varVM (r = 0.28, p < 0.05) and with smartphone mean VM (r = 0.05, p > 0.05). Daily ActiGraph MVPA showed no acceptable correlation with smartphone varVM (r = -0.12, p > 0.05) and with smartphone meanVM and (r = 0.32, p < 0.05).	Healthy participants: step counts/min and daily MVPA (not valid). Participants with multiple sclerosis: step counts/min and daily MVPA (not valid)
Studies conducted in laboratory conditions (n = 2)						
Ata et al.[27]	6-min walk test.	Step counts	NI	MPD, Pearson's correlation coefficient and Bland-Altman	Step counts showed acceptable mean difference (MPD = 5.7), acceptable correlation coefficient (r = 0.96 and no systematic differences (bias = 7.1, LoA = -34.3–46.5)	Valid
Rodriguez et al.[34]	10 minutes of walking at low speed, walking at high speed, going upstairs and downstairs, running, working at the office simulated in laboratory.	Count/min	Sasaki et al.[38]	ICC, Pearson's correlation coefficient, Bland-Altman and Kappa	Raw counts showed acceptable correlation coefficient (r = 0.94; ICC = 0.94) and no systematic differences (bias = 1.99, LoA = -5.8–5.7). Acceptable kappa (0.87) for light-, moderate-, vigorous- and very vigorous-intensity	Light-intensity, moderate-intensity, vigorous-intensity, and very vigorous-intensity (Valid)
Studies conducted in both free-living and laboratory conditions (n = 2)						
Hekler et al.[32]	Laboratory: 5 min lying down, sit and slouch, watching television, sweeping, walking on treadmill at 2 and 3 mph, jogging on treadmill at 5 mph, indoor cycling at 75 rpm, bicycling outside at 10 mph, standing. Free-living: 7 day of daily activities.	Counts/min	SB: Matthews et al.[37] and PA: Freedson et al.[36]	Pearson's correlation coefficient and Bland-Altman	Laboratory: acceptable correlation coefficient (r = 0.77 to 0.85). Free-living: acceptable absolute mean-level difference in sedentary (mean difference = -26 min/day) and MVPA (mean difference = -1.3 min/day) and no acceptable in light-intensity (mean difference = -111.2 min/day). No acceptable correlation for raw counts (r = 0.59, p < 0.05), sedentary (r = 0.44, p < 0.05), light-intensity PA (r = 0.38, p < 0.05) and MVPA (r = 0.67, p < 0.05). Systematic differences in sedentary (bias = -26.0, LoA = -279.5–227.6), light-intensity PA (bias = 111.3, LoA = -63.5–285.8) and no systematic differences in MVPA (bias = -1.3, LoA = -38.4–35.8)	Free-living: sedentary and MVPA (valid) and light-intensity (no valid). Laboratory (valid)
Maddison et al.[33]	Laboratory: 5 min resting and activity bouts at 4 km/h⁻¹, 6 km/h⁻¹, 10 km/h⁻¹, ≥12 km/h⁻¹. Free-living: 24 h.	kcal/min	Freedson et al.[36]	MPD, ICC, and Pearson's correlation coefficient	Free-living: EE showed no acceptable mean difference at light-intensity (MPD = 50.9), moderate-intensity (MPD = 129.7), vigorous-intensity (MPD = 87.0), very vigorous-intensity (MPD = 80.2) and total PA (MPD = 57.8). Acceptable correlation coefficient (r = 0.87, ICC = 0.83, p<0.05) was observed. Laboratory: EE showed no acceptable mean difference at rest (MPD = 77.6), 4 km/h (MPD = 13.6), 10 km/h (MPD = 14.8) and total PA (MPD = 13.9). EE showed acceptable difference at 6 km/h (MPD = 4.1) and 12 km/h (MPD = 8.2) and acceptable correlation and correlation coefficient (r = 0.92, ICC = 0.93, p<0.05).	Free-living: light-intensity, moderate-intensity, vigorous-intensity, very vigorous-intensity and total PA (not valid). Laboratory: 6 and 12 km/h (valid), rest, 4 and 10 km/h and total PA (not valid)

NI = no information; ICC = intraclass correlation coefficient; MPD = mean percentage difference; LoA = limits of agreement; EE = energy expenditure; METs = metabolic equivalents; MVPA = moderate-to-vigorous physical activity; meanVM = mean vector magnitude; varVM = mean variance of the vector magnitude.

A
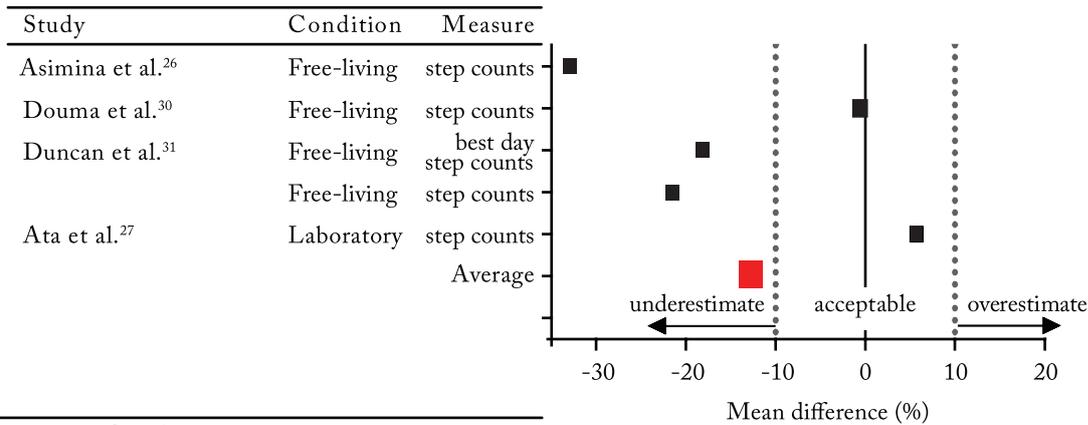
B
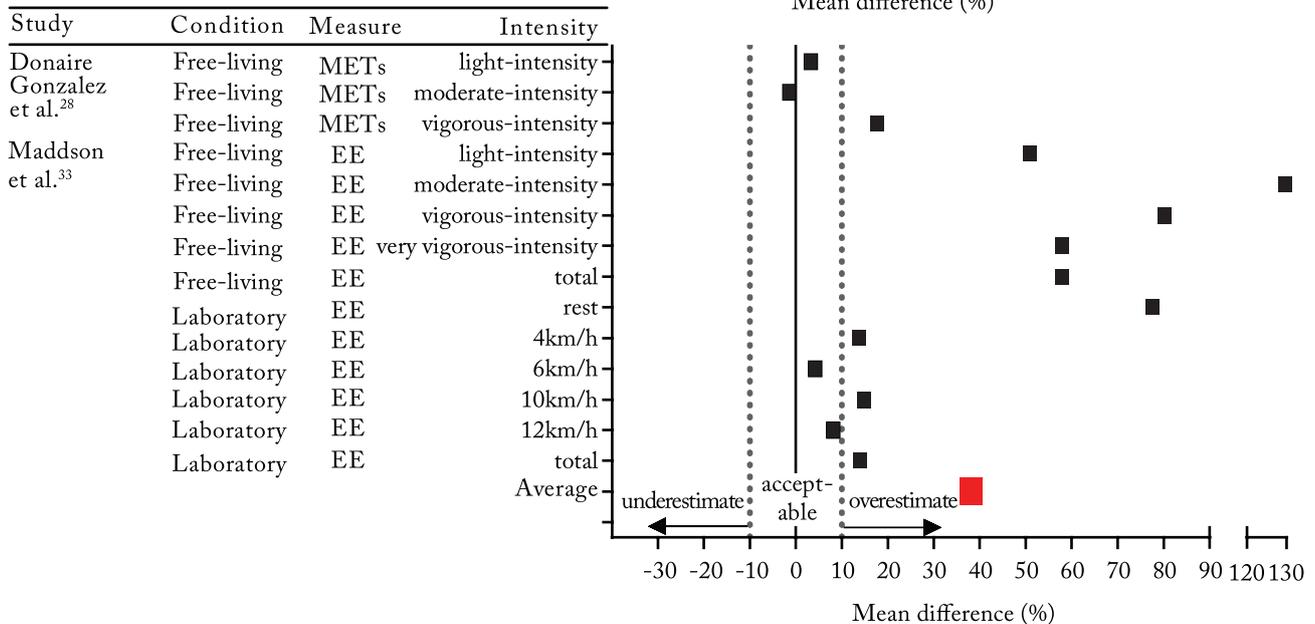
Figure 2 – MPD of the smartphone apps compared with the accelerometer for step counts (panel A) and intensity load (panel B). EE = energy expenditure; METs = metabolic equivalent.

-intensity PA. Another laboratory study[33] showed that the smartphone app overestimated intensity (MPD = 13.9) based on treadmill speed (i.e., km/h) - Table 3.

Figure 2 shows the MPD of the smartphone apps compared with the accelerometer (criterion) for step count measurements (panel A) based on four studies (five subgroups conducted in free-living and one in laboratory conditions) and according to intensity load (panel B) based on two studies (eight subgroups conducted in free-living conditions). For the studies that provided MPD or results from the equation adopted in this review, the smartphone apps underestimated step counts (average MPD = -12.7) and overestimated METs and energy expenditure according to intensity load (average MPD = 37.7). Both measures tended to have a mean difference that is closer to acceptable in laboratory con-

ditions. An overestimation of METs was also observed according to the increase in PA intensity (i.e., light-, moderate-, vigorous, and very vigorous-intensity).

Discussion

The current systematic review synthesized original validity studies that compared smartphone apps with a research-grade accelerometer to track PA or SB. The findings reinforced the notion of the controversial results regarding the criterion validity of smartphone apps for tracking PA[17], from which four were considered valid[27,29,30,34] and three were considered not valid[26,31,35] when examining all analyses of their respective designs; moreover, three studies were considered valid/not valid depending on the PA intensity (light-, moderate-, and vigorous-intensity) and conditions (free-living and la-

boratory conditions). The accuracy of the smartphone apps to track SB remains unclear owing to the scarcity of pertinent studies. Additionally, this review demonstrates that the studies involved a great variability of protocols (e.g., types of PA, all load components) that may justify the heterogeneity among the results.

The convenience sampling strategy used by all studies represents a limitation regarding the random effects found in randomized sampling strategies. The adoption of convenience sampling is understandable since this strategy is based on participants' accessibility and/or proximity to the laboratory[40]; however, the use of convenience sampling introduces a selection bias, impairing the external validity of the studies. Thus, it is not suitable for extrapolating the validation results of convenience sampling trials to the general population[41].

Failure to comply with the stipulated protocol (e.g., accelerometers wear and non-wear time) is responsible for the increase in the loss of data[42]. The studies included in this review covered only adult age groups; thus, the studies presented a limited possibility of extrapolating their findings for other populations, such as children or adolescents who demonstrate increased compliance difficulties because they may forget to wear the accelerometer, and may need additional strategies to overcome non-wear issues[43]. It is necessary to consider that these groups also have different cutoff values for the classification and recommendation of PA[3].

When seeking to introduce the use of smartphones for tracking PA or SB in epidemiologic research conducted in free-living conditions, the protocols may need to consider socially and contextually imposed behavior and practices. For example, all studies in this review were conducted in high-income countries where the violence rate is significantly lower as compared to low-income developing countries[44]. The Mobile Time website in partnership with Opinion Box revealed that in 2020, at least 100 million smartphones were stolen from 64 million of Brazilians over the age of 16[45]. However, the studies did not consider the rates of smartphones stolen in low-income countries. Thus, it is possible that due to being afraid to carry smartphones on their person (e.g., hand), people may not comply with the protocols of smartphone wearing properly. Hence, the inconsistency of data in free-living environments can be greater, since people usually prefer to carry smartphones where they feel safer, or even leave it at home. These issues should be further investigated.

It was observed that the condition and intensity of PA affect the validity of the app, in which apps provided low tracking accuracy for PA at vigorous intensities regarding different parameters (e.g., step counts and METs)[26,28]. Conversely, a previous systematic review[17] observed that smartphone apps were less accurate at lower PA intensities. Since smartphone apps use accelerometers that are built into smartphones to evaluate changes in interval acceleration of ± 2 g, underestimation or overestimation of measurements in high intensities of activity may occur due to the variability of triaxial movements between the locations where smartphones were carried (e.g., bag/backpack, pants pocket, or hands).

The current synthesis demonstrated the dependence of the accuracy of apps on the location of carrying a smartphone. A recent systematic review[17] showed that smartphones carried at the hip level provided the most accurate results among the studies. This may be due to the adoption of the hip-worn accelerometer protocol in all the studies. Since the apps use accelerometers that are built into smartphones for data collection, which usually represent the same or similar axis used by accelerometers (biaxial or triaxial), devices attached at the same location will suffer the same gravitational and dimensional oscillations. On the other hand, it is important to acknowledge that the research-grade accelerometer protocols have been migrating from the hip to the wrist recently[46,47] for the possibility of monitoring sleep outcomes, and this introduces a gap regarding the objective of the present review.

Overall, studies with laboratory designs showed a more robust pool of results, indicating better accuracy in monitoring PA compared to studies with free-living designs. In free-living designs, the lack of either load control or researcher presence, and wearing versus non-wearing of devices could result in higher measurement errors[48]. It is a challenge to epidemiological studies that seek to track PA and SB of populations on a large scale; thus, these findings reinforce the notion that future validation studies need to refine the study designs to validate low-cost methods (e.g., smartphone apps) to track the daily life behavior of the population.

The ActiGraph accelerometer models are among the most used research-grade accelerometers for such studies. Several studies have demonstrated ActiGraph accelerometers as a good tool for tracking PA or SB (attached at hip/wrist) in different subpopulations (e.g., children, adults, and the elderly)[49–52], conditions (e.g., free-living and laboratory)[50,53], and PA intensities (e.g., light-, moderate-, and vigorous-intensity)[49,54]. Howev-

er, a recent systematic review[55] observed unclear pieces of evidence regarding the accuracy of research-grade accelerometers, including ActiGraph models, to estimate the energy expenditure of the sample. Studies have shown that research-grade accelerometers overestimate energy expenditure during ambulation and stair climbing[55]. These findings show that research-grade accelerometer limitations may have been responsible for the conflicting results regarding energy expenditure among studies, and this statement is reinforced when we observe studies that have a high-quality assessment.

Only one study[32] investigated SB and demonstrated acceptable agreement between the accelerometer and smartphone app results. However, the results of this investigation may vary depending on the criteria established as SB measured by the accelerometer. Additionally, considering that smartphones may not be used or attached to people all time, it is possible that the pattern of smartphone allocation in real life can be very different from that adopted by participants during this free-living study; thus, it is unfeasible to measure such real-life behavior.

It was observed that a small number of studies[30–32] also conducted test-retest reliability analysis of the results. Recently, Lee and Hanage[56] argued the need for reproducibility in science, which suggests conducting future reliability tests aiming to develop robust software to track movement behaviors.

Finally, despite the digital revolution that happened approximately 30 years ago, there is a gap in the body of studies that validates smartphone apps as compared to the research-grade accelerometer for tracking PA or SB both regarding the date of the publication of the first research (since 2013) and the number of studies developed to date (n =10); thus, it is still necessary to advance knowledge in this field to include smartphone app data in scientific research that aims to measure PA and SB, which includes new studies that use accelerometers from other manufacturers (e.g., GENEActiv Original, Axivity AX3), attached to other locations (wrist), and new smartphone apps.

Limitations and Future perspectives

It was observed that studies focus on investigating mostly healthy adults and elderly populations. Thus, the results are not generalizable for children, adolescents, and adults with other diseases or disabilities. Additionally, the comparison of these results with lower-middle-income countries is limited because there are important sociocultural differences between these countries and high-income countries. Future validation studies addressing objective measurements of PA and SB should explore these issues.

The studies adopted different statistical methods to validate the apps, which makes it difficult to compare the results and conduct more rigorous quantitative analysis (e.g., meta-analysis). Additionally, the MPD analysis conducted in this review needs to be considered limited because it did not consider the standard deviation and sample size of the groups.

Six apps considered in this review were unavailable for downloading, three of which were developed with models developed by the studies (Supplementary Data 2), which creates a limitation for reproducibility and the development of new studies.

Since the validation of smartphone apps to track PA and SB is an interesting strategy, as it has a lower cost and a greater number of smartphone users, allowing its use in epidemiological studies is necessary, but it is also important to consider the efficiency of the app, which reflects in the validation of app for the general population and not small sample (e.g., 50, 100 people). Thus, we strongly recommend that future studies recruit a larger number of participants to increase the statistical power of a study and also adopt a randomized sampling strategy.

Recently, Community Mobility Reports[57], a dataset provided by Google from their mobile device apps showed changes in mobility before and during the current COVID-19 pandemic. Google indicates that this dataset can be used for public policy strategies to confront the pandemic; on the other hand, in future studies and public policy strategies, insights should be discovered for the use of this strategy (Google mobility track) to track PA and to combat another pandemic, that of SB.

This systematic review showed that four smartphone apps were considered valid for tracking PA, but six were not valid or fully valid. Although smartphone apps are slightly more accurate for tracking PA in laboratory designs when compared with free-living designs, there is still little evidence to validate apps that monitor PA. Overall, PA intensity and the location where the smartphone is carried or attached tend to affect the accuracy of the PA measures. The accuracy of smartphone apps for measuring SB as compared to research-grade accelerometers is unclear because of the lack of original studies. Despite being practical and cheaper, smartphone apps are not fully valid for mass use in PA and SB epidemiological surveillance. We recommend that

when using smartphone apps for monitoring PA, caution is necessary, considering that the apps need to improve the accuracy of their measurements.

Conflict of interest

The authors declare no conflict of interest.

Author's contributions

Prado RCR, Knebel MTG, Ribeiro EHC, Teixeira IP, Sasaki JE, Araujo LV, Guerra PH and Florindo AA participated in the design of the manuscript. Prado RCR, Knebel MTG and Florindo AA were responsible for the literature search, data collection, data analysis and interpretation, and manuscript writing. Prado RCR, Knebel MTG, Ribeiro EHC, Teixeira IP, Sasaki JE, Araujo LV, Guerra PH and Florindo AA performed a critical and final review of the content.

References

1. Marques A, Santos T, Martins J, Matos MG, Valeiro MG. The association between physical activity and chronic diseases in European adults. Eur J Sport Sci. 2018;18(1):140-49.
2. Lee IM, Skerrett PJ. Physical activity and all-cause mortality: what is the dose-response relation?. Med Sci Sports Exerc. 2001;33(6 Suppl):S459-S94.
3. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451-62.
4. Jones RA, Hinkley T, Okely AD, Salmon J. Tracking physical activity and sedentary behavior in childhood: a systematic review. Am J Prev Med. 2013;44(6):651-58.
5. Hayes G, Dowd KP, MacDonncha C, Donnelly AE. Tracking of Physical Activity and Sedentary Behavior From Adolescence to Young Adulthood: A Systematic Literature Review. J Adolesc Health. 2019;65(4):446-54.
6. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act. 2012;9:103.
7. Chastin SFM, Dontje ML, Skelton DA, Čukić I, Shaw RJ, Gill JMR et al. Systematic comparative validation of self-report measures of sedentary time against an objective measure of postural sitting (activPAL). Int J Behav Nutr Phys Act. 2018;15(1):21.
8. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical guide to measuring physical activity. J Acad Nutr Diet. 2014;114(2):199-208.
9. van der Ploeg HP, Merom D, Chau JY, Bittman M, Trost SG, Bauman AE. Advances in Population Surveillance for Physical Activity and Sedentary Behavior: Reliability and Validity of Time Use Surveys. Am J Epidemiol. 2010;172(10):1199-206.
10. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014;48(13):1019-23.
11. Wong TC, Webster JG, Montoye HJ, Washburn R. Portable Accelerometer Device for Measuring Human Energy Expenditure. IEEE Trans Biomed Eng. 1981;BME-28(6):67-471.
12. John D, Freedson P. ActiGraph and Actical physical activity monitors: a peek under the hood. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S86-9.
13. Matthews CE, Hagstromer M, Pober DM, Bowles HR. Best Practices for Using Physical Activity Monitors in Popula-tion-Based Research. Med Sci Sports Exerc. 2012;44(1):S68-S76.
14. Fitbit Reports $571M Q4'17 and $1.616B FY'17 Revenue. fitbit.com. https://www.webcitation.org/6xuzBvDVV (accessed Mar 14, 2021).
15. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat A, Yoo JY et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth. 2018;6(8):e10527.
16. BankMyCell. How many smartphones are in the world? https://www.bankmycell.com/blog/how-many-phones-are-in-the-world (accessed October 20, 2020).
17. Silva AG, Simões P, Queirós A, Rodrigues M, Rocha NP. Mobile Apps to Quantify Aspects of Physical Activity: A Sys-tematic Review on its Reliability and Validity. J Med Syst. 2020;44(2):51.
18. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for system-atic reviews of interventions. 2nd ed. Chichester: Wiley; 2019
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TM, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
20. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. Lancet Infect Dis. 2010;10(4):226.
21. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
22. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. J Eval Clin Pract. 2012;18(6):1126-32.
23. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J. 2015;14:40.
24. Aaronson N, Alonso J, Burnam A, Lohr NK, Patrick DL, Perrin E et al. Assessing health status and quality-of-life in-struments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205.
25. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008;5:56.
26. Asimina S, Chapizanis D, Karakitsios S, Kontoroupis P, Asimakopoulos DN, Maggos T et al. Assessing and enhancing the utility of low-cost activity and location sensors for exposure studies. Environ Monit Assess. 2018;190(3):155.

27. Ata R, Gandhi N, Rasmussen H, El-Gabalawy O, Gutierrez S, Ahmad A et al. Clinical validation of smartphone-based activity tracking in peripheral artery disease patients. NPJ Digit Med. 2018;1.

28. Donaire-Gonzalez D, de Nazelle A, Seto E, Mendez M, Nieuwenhuijsen MJ, Jerrett M. Comparison of physical activity measures using mobile phone-based CalFit and Actigraph. J Med Internet Res. 2013;15(6):e111.

29. Donaire-Gonzalez D, Valentin A, van Nunen E, Curto A, Rodrigues A, Fernandez-Nieto M et al. ExpoApp: An integrat-ed system to assess multiple personal environmental exposures. Environ Int. 2019;126:494-503.

30. Douma JAJ, Verheul HMW, Buffart LM. Feasibility, validity and reliability of objective smartphone measurements of physical activity and fitness in patients with cancer. BMC Cancer. 2018;18(1):1052.

31. Duncan MJ, 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(15):1695-704.

32. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. JMIR mHealth uHealth. 2015;3(2):e36.

33. Maddison R, Gemming L, Monedero J, Bolger L, Belton S, Issartel J et al. Quantifying Human Movement Using the Movn Smartphone App: Validation and Field Study. JMIR Mhealth Uhealth. 2017;5(8):e122.

34. Rodriguez VH, Medrano C, Plaza I, Corella C, Abarca A, Julian JA. Comparison of Several Algorithms to Estimate Ac-tivity Counts with Smartphones as an Indication of Physical Activity Level. IRBM. 2019;40(2):95-102.

35. Zhai Y, Nasseri N, Poettgen J, Gezhelbash E, Heesen C, Stellmann J-P. Smartphone Accelerometry: A Smart and Reli-able Measurement of Real-Life Physical Activity in Multiple Sclerosis and Healthy Individuals. Front Neurol. 2020;11:688.

36. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998;30(5):777-81.

37. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol. 2008;167(7):875-81.

38. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. 2011;14(5):411-16.

39. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DRJ. Refined two-regression model for the ActiGraph accelero-meter. Med Sci Sports Exerc. 2010;42(5):1029-37.

40. Bornstein MH, Jager J, Putnick DL. Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Stan-dards. Dev Rev. 2013;33(4):357-70.

41. Hagströmer M, Ainsworth BE, Kwak L, Bowles HR. A checklist for evaluating the methodological quality of validation studies on self-report instruments for physical activity and sedentary behavior. J Phys Act Health. 2012;9 Suppl 1:S29-S36.

42. Perry MA, Hendrick PA, Hale L, Baxter GD, Milosavljevic S, Dean SG et al. Utility of the RT3 triaxial accelerometer in free living: An investigation of adherence and data loss. Appl Ergon. 2010;41(3):469-76.

43. Belton S, O'Brien W, Wickel EE, Issartel J. Patterns of noncompliance in adolescent field-based accelerometer research. J Phys Act Health. 2013;10(8):1181-5.

44. Wolf A, Gray R, Fazel S. Violence as a public health problem: an ecological study of 169 countries. Soc Sci Med. 2014;104(100):220-27.

45. Mobile Time. Mais de 100 milhões de celulares já foram roubados ou furtados no Brasil. https://www.mobiletime. com.br/noticias/23/07/2020/mais-de-100-milhoes-de-celulares-ja-foram-roubados-ou-furtados-no-brasil/ (accessed Apr 14, 2021).

46. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M et al. A Novel, Open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One. 2015;10(11):e0142533.

47. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. J Sci Med Sport. 2017;20(1):75-80.

48. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. 2015;12:159.

49. Santos-Lozano A, Santín-Medeiros F, Cardon G, Torres-Luque G, Bailón R, Bergmeir C et al. Actigraph GT3X: vali-dation and determination of physical activity intensity cut points. Int J Sports Med. 2013;34(11):975-82.

50. Kelly LA, McMillan DG, Anderson A, Fippinger M, Fillerup G, Rider J. Validity of actigraphs uniaxial and triaxial ac-celerometers for assessment of physical activity in adults in laboratory conditions. BMC Med Phys. 2013;13(1):5.

51. Steenbock B, Wright MN, Wirsik N, Brandes M. Accelerometry-Based Prediction of Energy Expenditure in Preschoo-lers. J Meas Phys Behav. 2(2):94-102.

52. Hall KS, Howe CA, Rana SR, Martin CL, Morey MC. METs and accelerometry of walking in older adults: standard versus measured energy cost. Med Sci Sports Exerc. 2013;45(3):574-82.

53. McClain JJ, Sisson SB, Tudor-Locke C. Actigraph accelerometer interinstrument reliability during free-living in adults. Med Sci Sports Exerc. 2007;39(9):1509-14.

54. Vanhelst J, Béghin L, Turck D, Gottrand F. New validated thresholds for various intensities of physical activity in adoles-cents using the Actigraph accelerometer. Int J Rehabil Res. 2011;34(2):175-7.

55. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med. 2020;54(6):332-40.

56. Lee RS, Hanage WP. Reproducibility in science: important or incremental? The Lancet Microbe. 2020;1(2):e59-60.
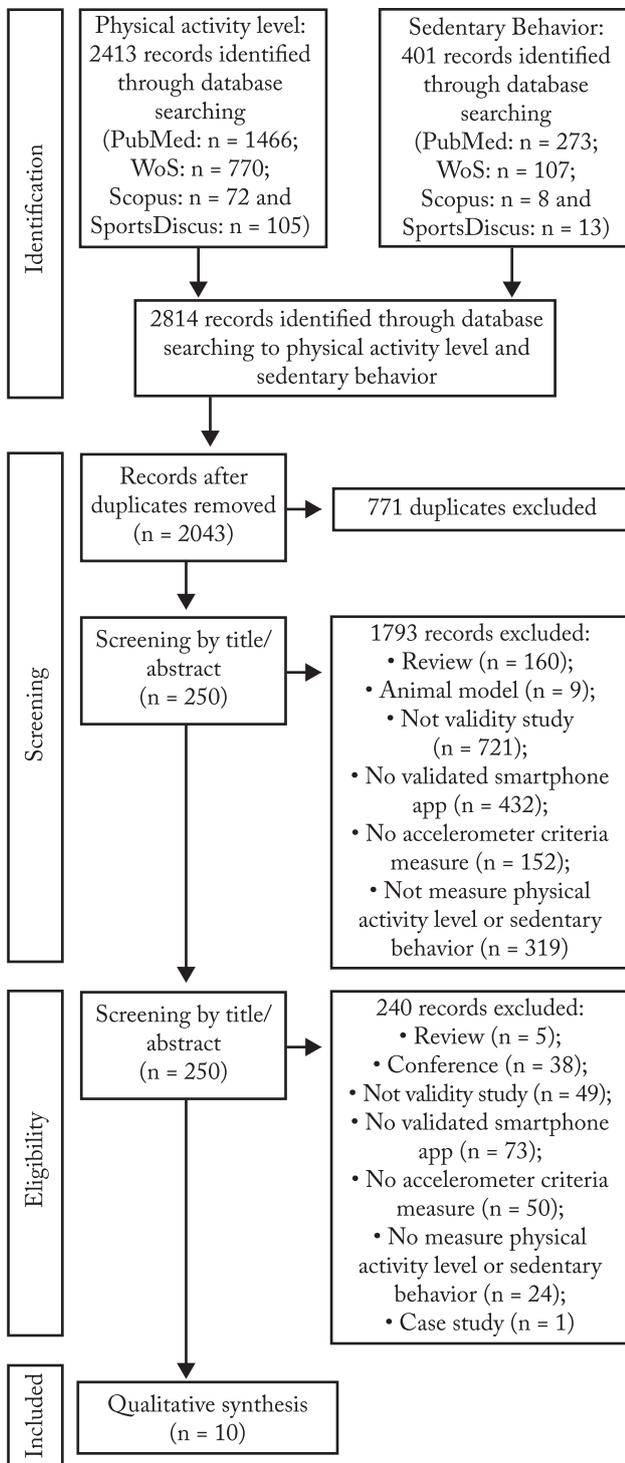
57. Google. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility?hl=en (accessed Apr 21, 2021).

Supplementary Data

Supplementary Data 1 – PRISMA flowchart of eligibility criteria and study selection.

Supplementary Data 2 – URL for download smartphone apps

Authors (year)	App name	Link Provide by corresponding author
Asimina et al.[26]	Moves	unavailable
Ata et al.[27]	VascTrac	https://apps.apple.com/us/app/vasctrac/id1121791155
Donaire-Gonzalez et al.[28]	CalFit	unavailable
Donaire-Gonzalez et al.[29]	Expo	unavailable
Douma et al.[30]	Walkmeter	https://apps.apple.com/nl/app/walkmeter-walking-hiking-gps/id330594424
Duncan et al.[31]	Health	https://www.apple.com/ca/ios/health/
Hekler et al.[32]	Custom app	unavailable
Maddison et al. [33]	Movn	iOS: https://apps.apple.com/us/app/movn-rehab/id1191991344 Android: https://play.google.com/store/apps/details?id=com.movn_react_native
Rodriguez et al.[34]	Custom app	unavailable
Zhai et al.[35]	Custom app	unavailable