BMJ Open Sport & Exercise Medicine

Assessment of 24-hour physical behaviour in children and adolescents via wearables: a systematic review of free-living validation studies

Marco Giurgiu ,[1,2] Simon Kolb,[1] Carina Nigg,[1,3] Alexander Burchartz,[1] Irina Timm,[1] Marlissa Becker,[4] Ellen Rulf,[1] Ann-Kathrin Doster,[1] Elena Koch,[1] Johannes B J Bussmann,[5] Claudio Nigg ,[6] Ulrich W Ebner-Priemer,[2,7] Alexander Woll[1]

To cite: Giurgiu M, Kolb S, Nigg C, *et al*. Assessment of 24-hour physical behaviour in children and adolescents via wearables: a systematic review of free-living validation studies. *BMJ Open Sport & Exercise Medicine* 2022;**8**:e001267. doi:10.1136/bmjsem-2021-001267

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjsem-2021-001267).

Accepted 19 April 2022

For numbered affiliations see end of article.

Correspondence to
Dr Marco Giurgiu;
marco.giurgiu@kit.edu

ABSTRACT

Objectives Studies that assess all three dimensions of the integrative 24-hour physical behaviour (PB) construct, namely, intensity, posture/activity type and biological state, are on the rise. However, reviews on validation studies that cover intensity, posture/activity type and biological state assessed via wearables are missing.

Design Systematic review. The risk of bias was evaluated by using the QUADAS-2 tool with nine signalling questions separated into four domains (ie, patient selection/study design, index measure, criterion measure, flow and time).

Data sources Peer-reviewed validation studies from electronic databases as well as backward and forward citation searches (1970–July 2021).

Eligibility criteria for selecting studies Wearable validation studies with children and adolescents (age <18 years). Required indicators: (1) study protocol must include real-life conditions; (2) validated device outcome must belong to one dimension of the 24-hour PB construct; (3) the study protocol must include a criterion measure; (4) study results must be published in peer-reviewed English language journals.

Results Out of 13 285 unique search results, 76 articles with 51 different wearables were included and reviewed. Most studies (68.4%) validated an intensity measure outcome such as energy expenditure, but only 15.9% of studies validated biological state outcomes, while 15.8% of studies validated posture/activity type outcomes. We identified six wearables that had been used to validate outcomes from two different dimensions and only two wearables (ie, ActiGraph GT1M and ActiGraph GT3X+) that validated outcomes from all three dimensions. The percentage of studies meeting a given quality criterion ranged from 44.7% to 92.1%. Only 18 studies were classified as 'low risk' or 'some concerns'.

Summary Validation studies on biological state and posture/activity outcomes are rare in children and adolescents. Most studies did not meet published quality principles. Standardised protocols embedded in a validation framework are needed.

PROSPERO registration number CRD42021230894.

KEY MESSAGES

WHAT IS ALREADY KNOWN

⇒ There is a rising interest in using wearables to assess all three dimensions of the integrative 24-hour physical behaviour construct, namely, intensity, posture/activity type and biological state.
⇒ Recently, generic validation frameworks for wearables integrating various conditions have been proposed.
⇒ Reviews on studies investigating the validity of wearables under free-living conditions covering the full integrative 24-hour physical behavioural construct are missing.

WHAT ARE THE NEW FINDINGS

⇒ Most studies validated intensity outcomes such as energy expenditure, but validation studies on biological states (such as asleep or awake) and posture/activity type outcomes are rare in children and adolescents.
⇒ Most reviewed validation studies do not meet currently published core principles regarding study quality. Notably, no reviewed study that validated a biological state or posture/activity type outcome was classified as 'low risk'.
⇒ Thirty of the 51 included wearables were validated only once. No wearable was identified that measures all dimensions with moderate to strong validity.

INTRODUCTION

Within the last two decades, the forefront of activity research moved from assessing single parameters such as steps or counts over the assessment of sedentary behaviour (SB) and physical activity (PA) in parallel to an integrated perspective of different movement and non-movement patterns, the so-called 24-hour activity cycle (24-HAC).[1 2] This development was pushed by empirical evidence that those different parameters contribute independently to health and resulted in WHO

guidelines[3] that provided separate recommendations for specific behaviours. For example, current 24-hour movement guidelines for children and youth[4] emphasise that being regularly physically active, reducing sedentary time and having a healthy sleep pattern at young ages will contribute to physical and mental health, reduce the risk of developing obesity in childhood and are associated with non-communicable diseases later in life.[5 6] The gradual shift from focusing on a single behaviour such as PA to a multiperspective focus on 24-hour physical behaviour (PB) (ie, including sleep, SB and PA) has also been theoretically addressed. Rosenberger *et al*[1] introduced the 24-HAC model as a new paradigm for PA, and Tremblay *et al*[2] provided a conceptual model of movement-based terminology around the 24-hour cycle. In particular, the 24-hour PB cycle comprises three movement and non-movement behaviours (ie, PA, SB and sleep; see online supplemental appendix 1). Each of the three behaviours can be differentiated by specific characteristics. For example, the sleep state is characterised by reduced or absent consciousness, whereas PA and SB appear while being awake. The Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS) extended the approach by subdividing the 24-hour PB construct into three behaviours and applying different dimensions.[7] In detail, each behaviour covers aspects of biological (ie, asleep or awake), posture (eg, lying, sitting or upright) and intensity (eg, light, moderate or vigorous) dimensions. For example, SB is defined as any waking behaviour (ie, biological state) characterised by an energy expenditure of ≤1.5 metabolic equivalents (ie, intensity) while in a sitting, reclining or lying posture (ie, posture).[2] Thus, the differentiation between PA, SB and sleep requires a valid and simultaneous assessment of all three dimensions.

Technical developments over the past decades have led to the opportunity to use wearables (eg, research-grade and consumer-grade accelerometers or pedometers) to capture high-frequency data about human movement and non-movement behaviour in daily life over longer periods.[8 9] Since 2016, wearable technology has been a leading fitness trend with an estimated $100 billion industry.[10] In summary, there is a growing commercial industry of wearable technology, a growing number of research studies that integrated device-based methods to capture PB data and discussions about whether it is 'prime time' for scientifically validated wearables to be global PB surveillance methodologies.[11 12] However, applying wearables in health studies offers methodological and practical challenges such as data processing, monitoring protocols or quality criteria such as validity[9] while aiming to allow for valid interstudy comparisons.

The concept of validity is a fundamental criterion to evaluate the quality of an instrument, referring to the degree to which it truly measures the construct it purports to measure.[13] Researchers interested in the 24-hour PB cycle are most interested in criterion-referenced validity because the parameters that they are attempting to measure are highly objective.[14] In line with the wearable technology's popularity in health research, the number of validation studies has increased dramatically over the past decades. Recently, researchers emphasise the importance of performing standardised validation procedures.[15–17] For example, Sperlich and Holmberg[17] strongly recommend controlling and monitoring the launching of wearable technology for health and fitness purposes and call for independent scientific validation procedures in terms of 'evidence-based marketing claims'.[17]

Collaborations such as the INTERLIVE network started developing standardised protocols to validate consumer wearables for steps[15] and heart rate.[18] As a broader framework, Keadle *et al*[16] introduced a stage process of validity to facilitate the development and validation of processing methods to assess PB via wearables.[16] The framework contains five validation phases with increasing levels, starting from device manufacturing and culminating with application in health studies. After mechanical (phase 0) and calibration testing (phase 1), validation studies are suggested with a fixed and semistructured evaluation under laboratory (phase II) as well as under real-life conditions (phase III) in which participants can complete their natural everyday behaviour.[16] Optimally, the validation of devices occurs through all stages before applying the device in health research studies (phase IV). Since error rates differ between laboratory and real-life conditions,[15] it is advisable to capture a wide array of activities in daily living under real-life conditions. Moreover, under laboratory conditions, some researchers may instruct participants to perform specific activities, which can lead to unnaturally performed activities (eg, the Hawthorne effect or reactivity bias).[19] Therefore, it is crucial to quantify measurement error in an unconstrained free-living environment and compare wearables' outcomes with a reference measure such as video recordings or the doubly labelled water method. The realisation of standardised validation protocols embedded in a framework[15 16] are helpful to inform consumers and can aid researchers in study design when selecting the appropriate wearable.[20 21] Transparency of the results may foster manufacturers' innovation to achieve improved validity and inform practitioners while incorporating wearables into daily clinical practice.[15]

Research purpose

Given the rapid increase and availability of research and consumer-grade wearables, a free-living validation is required to ensure appropriate conclusions for surveillance, epidemiological and intervention studies. Although previous reviews focused on the issue of validity (eg, intensity levels of PA and SB,[22] energy expenditure,[23 24] steps[20] and sleep-related outcomes[25–27]); however, this field has not been assessed to date from a comprehensive 24-hour PB perspective. Following the age-dependent classification of WHO guidelines,[28] we selected both children and adolescents. The 24-hour PB construct is an important aspect of physical and mental development

across the paediatric age range.[28 29] Moreover, movement patterns of children and adolescents are specific (eg, intermittent and sporadic[30 31]), which may affect the initialisation of devices when assessing 24-hour PB. Finally, this review focuses on the following purposes: first, as our main purpose, we would like to raise researchers' and consumers' attention to the quality of published validation protocols while aiming to identify and compare specific consistencies/inconsistencies between validation protocols. To evaluate the quality of the studies, we followed core principles, recommendations and expert statements[14–16 32] with published quality criteria (eg, study duration, number of included participants, selection of criterion measure and data synchronisation). Second, we would like to provide a comprehensive and historical overview on which wearable has been validated for which purpose, and whether they show promise or not for being used in further studies.

METHODS

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses reporting guidelines (checklist: see online supplemental appendix 2).[33]

Search strategy and study selection

We used a search string that included terms for (1) examine validity, (2) device/type of wearable and (3) dimensions of the 24-hour PB construct. An a priori pilot search was conducted to optimise the final term (see online supplemental appendix 3). Publications were searched from 1970 to December 2020 using the following databases: EbscoHost, IEEE Xplore, PubMed, Scopus and Web of Science. We reran the search in July 2021 to check for updates and checked the reference lists of included studies for publications that may meet the inclusion criteria.

All articles were imported to a Citavi library (Citavi V.6.8, Swiss Academic Software GmbH, Swiss). After removing all duplicates, the study selection process included three screening phases on eligibility. In phase I, two reviewers screened the titles of the publications independently (MG and ER) while focusing on the selected key terms of the search string (ie, validity, wearable and parameter of 24-hour PB construct). Articles were excluded only if both reviewers categorised an article as not eligible for review purposes. In the second phase, two reviewers screened and reviewed the abstracts of the publications independently (MG and ER). Discrepancies in screening were resolved by consulting a third reviewer (EK). Finally, in the third phase, the full texts of the remaining articles were assessed for eligibility by seven members of the author's team (MG, ER, AK-D, SK, AB, IT and CaN). Each article was screened independently by at least two reviewers. Discrepancies in screening were resolved by discussion until consensus was reached. Reviewers were not blinded to author or journal information (see online supplemental appendix 4).

Inclusion and exclusion criteria

Based on the PICO principle,[34] we included peer-reviewed, English-language publications that met the following criteria:

1. *Population*: Participants were children and adolescents <18 years regardless of health conditions.
2. *Intervention*: Any wearable validation study in which at least one part of the study was conducted under free-living (naturalistic/real-life) conditions (eg, at participants' homes or school and without instructions on when to start or stop a particular activity).
3. *Control/comparison*: Studies were included only if they described a criterion measure.
4. *Outcomes*: Studies were included in which the wearable outcome(s) could be classified into at least one dimension of the 24-hour PB construct (ie, biological state, posture/activity type or intensity[7]; see online supplemental appendix 1). Although the constructs' type of posture (eg, lying, sitting or upright) and type of activity (eg, descriptions of body movements, such as walking or cycling, as well as of specific functional activities, such as cooking or reading) describe different aspects of 24-hour PB, we have combined them into one section, since the output of some devices provides combined parameters of postural and activity type of information.

Data extraction

Data were independently extracted by two authors (MG, ER, AK-D, SK, AB, IT or CN). Discrepancies were discussed until consensus was reached. The following study details were extracted: author; year; location; population information (sample size, mean age of participants, percentage of women and ethnicity); measurement period; validated wearable (wearing position, software, epoch length and algorithm/cut-point); dimension of the 24-hour PB construct; validated outcome; criterion measure; statistical analyses for validation purposes; conclusion and funding information.

Data synthesis

Given the wide range of different study protocols in terms of varying conditions (eg, wear location, measurement duration, sample size, statistical analyses or criterion measure), we conducted a narrative synthesis based on the reported results/conclusions. The data synthesis focused on our secondary purpose, that is, whether the included wearables show promise or not for being used in further studies. In particular, we classified the studies as ↑ (ie, moderate to strong validity), ↔ (ie, mixed results) and ↓ (ie, poor or weak validity). Each article was classified independently by at least two reviewers. Discrepancies in classification were resolved by discussion until consensus was reached.

Quality assessment

According to our main purpose, that is, raising researchers' and consumers' attention to the quality

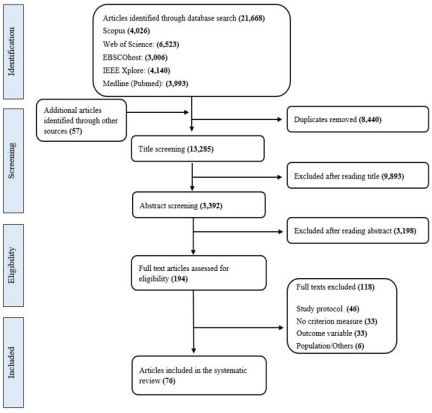

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart illustrating the literature search and screening process.

of published validation protocols, we conducted a quality assessment and evaluated the risk of bias. Each article was evaluated using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[32] The tool is composed of four different domains (ie, patient selection, index measure, criterion measure and flow/timing). Following the QUADAS-2 guidelines, we selected a set of signalling questions for each domain and added questions modified from the QUADAS-2 background document based on core principles, recommendations and expert statements for validation studies[14–16 32] (see table 3). The risk of bias assessment was conducted independently by at least two authors. Discrepancies were discussed until consensus was reached. The study quality was evaluated at the domain level; that is, if all signalling questions for a domain were answered 'yes', then the risk of bias was deemed to be 'low'. If any signalling question was answered 'no', then the risk of bias was deemed to be 'high'. The 'unclear' category was only used when insufficient data were reported for evaluation. Based on the domain-level ratings, we created a decision tree to evaluate the overall study quality as 'low risk', 'some concerns' or 'high risk' (see online supplemental appendix 5).

RESULTS

The search resulted in 13 285 unique records, with 76 publications (representing 74 unique studies; see figure 1) being included.[35–110] In particular, 68.4% (n=52) of all studies were classified into the intensity dimension, 15.8% (n=12) into the posture/activity type dimension and 15.8% (n=12) into the biological state dimension. None of the included studies validated outcomes from two different dimensions; that is, no study validated intensity and postural/activity type outcomes at the same time.

Participant and study characteristics

Of the studies included, 71.1% were published within the last decade (≥2011), indicating the increasing use of wearable technologies for PB measurement (see table 1); 88.2% (n=67) were conducted in wealthier high-income countries from North America, Europe or Australia/Oceania. The number of participants ranged between 3 and 225, while most studies (57.9%, n=44) recruited between 21 and 50 participants. The mean age of the participants' samples ranged between newborn infants (median 37 weeks) and older adolescents (17±1 years). In most studies, the mean age of the sample was between 8 and 13 years. Healthy participants were recruited in 97.4% (n=74) of all studies. One study recruited children with congenital heart disease, and one study recruited both healthy children and children with autism. Participants' ethnicity was reported in 27.6% (n=21) of all studies. The measurement duration of the reviewed study protocols varied between approximately 30 min and up to 14 days. For example, studies that focused on biological state or posture/activity type outcomes reported in 14 of 24 studies a duration of ≤1 day. The majority of studies (88.2%, n=67) conducted statistical analyses at the person/study level (eg, correlations, t-tests and repeated measures analysis of variance (ANOVA)). Five studies conducted both person-level/study-level analyses as well as epoch-by-epoch comparisons (eg, sensitivity and specificity). Two studies used machine-learning approaches to identify activity types. Six studies reported that the manufacturer was involved in study funding or loaned the devices, or one of the authors declared a relation to the company of the validated wearable. In 44.7% (n=34) of all studies, funding was independent of the manufacturer, and the authors declared no conflict of interest. In 15.8% (n=12) of all studies, neither information about funding nor any information about conflict of interests was reported, whereas in the remaining 24 studies, at least funding information or conflict of interest statement was reported and without any relation to the manufacturer. Detailed data extraction is reported as a supplement (see online supplemental appendix 6).

Wearables

We identified 51 different wearables, of which 31 were classified as research and 20 as consumer-grade devices. The type of wearables varied across uniaxial, biaxial or triaxial accelerometers and pedometers. Detailed technical information for each wearable is available as a supplement (see online supplemental appendix 7). Twenty-two studies included multiple sensors or wearing positions to enable comparison between different devices or wearing locations. The variation of brands within a study protocol ranged from one to four. In particular, 76.3% (n=58) of all studies included one brand of wearable, whereas 23.7% (n=18) included two to four different brands of wearables. We identified 10 different validated outcomes (see table 2) The hip/waist and wrist positions were most often used for validation purposes. In 53.9% (n=41) of all studies, the authors provided information about the software application used for data preprocessing issues.

Table 1 Summary of data extraction: participant and study characteristics

Category		Total (N=76)	Biological state (n=12)	Posture/activity type (n=12)	Intensity (n=52)
Publication year	≤1999	4		1	3
	2000–2010	18	1		17
	≥2011	54	11	11	32
Study location	Africa	1			1
	Asia	7	1	1	5
	Europe	29	1	8	20
	North America	31	9	1	21
	Australia/Oceania	7	1	2	4
	South America	1			1
Participants (N)*	≤19	14	2	3	9
	20–50	46	9	8	29
	≥51	17	1	1	15
Mean age (years)†	0–3.9	12	4	2	6
	4.0–7.9	16	1	3	12
	8.0–12.9	37	6	5	26
	13–18	9	2		7
Sex (%, female)‡	0–25	3		1	2
	26–74	68	12	9	47
	75–100	2			2
Measurement duration (days)	≤1	34	7	7	20
	2–6	15	2	2	11
	≥7	27	3	3	21
Criterion measure	Doubly labelled water	14			14
	Heart telemetry	1			1
	Indirect calorimetry	4			4
	Observation (direct)	11			11
	Observation (video)	11	2	8	1
	Polysomnography	6	6		
	Questionnaire/diary	3	2		1
	Wearable	26	2	4	20
Statistical analyses	Epoch-by-epoch	14	6	3	5
	Person/study level	67	8	8	51

*One article reported two studies and was counted twice.
†One study reported two subgroups, and the subgroups were counted twice; three studies were not included in the summary statistics due to the lack of age information.
‡Three studies were not included in the summary statistics due to the lack of sex information.

Across all studies, the selected epoch length varied from 1 s to 1 min, while 15 s and 1 min were most often reported. In 44.7% (n=34) of all studies, some information about the used algorithm, equation or cut-points was reported.

Study quality

In total, we included nine signalling questions as quality criteria to evaluate the risk of bias. The percentage of meeting a given criteria ranged from 44.7% to 92.1% (see table 3). On average, 5.7 of 9 questions were answered with yes (ie, meeting the criteria). Studies validating a biological state, intensity or posture/activity type outcome met on average 6.0, 5.8 and 5.3 out of 9 questions with yes, respectively. In 46.1% (n=35) of all studies, the reference standard was in line with the suggested criterion measures.[16] Wearables were the most frequently selected criterion reference in 34.2% (n=26) of the studies (see table 2). Based on our classification tree to evaluate the overall study quality (see online supplemental appendices 5 and 8), seven studies were classified as low risk (all validated an intensity outcome). Furthermore, 11 studies were classified with some concerns and 58 studies with

Table 2 Summary of data extraction: wearables

Category		Total (N=76)	Biological state (n=12)	Posture/activity type (n=12)	Intensity (n=52)
Type	Uniaxial accelerometer	33	5	2	26
	Biaxial accelerometer	8	3		5
	Triaxial accelerometer	47	5	14	28
	Pedometer	9			9
Outcome	Sleep time	8	8		
	Sleep–wake metrics	4	4		
	Different postures/types	8		8	
	Time in SB	16		4	12
	Time in LPA	10			10
	Time in MVPA	19			19
	Energy expenditure	21			21
	Steps	16			16
	Counts	10		1	9
Wear position*†	Ankle	5	4		1
	Chest	2			2
	Hip/waist	55	5	7	43
	Lower back	2			2
	Thigh	4		4	
	Upper arm	5			5
	Wrist	30	9	7	14
	Backpack/pockets	3		2	1

*One study did not report any information about the sensor wearing position.
†If studies included multiple devices or different wearing positions, we counted each wearing position.

high risk. Figure 2 illustrates the overall study quality on a study level separated by each dimension of the 24-hour PB construct.

Validity

Across all studies (n=76), we classified 102 validation results of 51 different wearables. In particular, we ranked 43.14% (n=44) results/conclusions as '↑' (ie, moderate to strong validity), 38.23% (n=39) as '↔' (ie, mixed validity) and 18.63% (n=19) as '↓' (ie, poor or weak validity). Table 4 provides an overview for each wearable separated by different age groups. Of those 51 different wearables, 58.8% (n=30) were validated once; 19.6% (n=10) were validated in two different studies; 11.8% (n=6) were validated in three different studies; and 9.8% (n=5) were validated in more than three different studies. ActiGraph GT3X+ (n=10), ActiGraph GT1M (n=7), Yamax Digiwalker SW-200 (n=6), ActiGraph AM7164 (n=5) and Tritrac R3D (n=4) were used most often in the included validation studies. Most wearables (n=42) had been used for the validation of only one dimension of the 24-hour PB construct. In particular, 33 wearables have been used only for the validation of intensity outcomes, whereas five wearables for the validation of posture/activity type and four wearables for the validation of biological state outcomes. In contrast, we identified three wearables (ie, Actiwatch spectrum, Actiwatch-L and Fitbit Charge HR) that validated both intensity and biological state outcomes and three wearables (ie, Actical, ActiGraph GT3X and GENEActiv) that validated both intensity and posture/activity types of outcomes. Moreover, two wearables (ie, ActiGraph GT1M and ActiGraph GT3X+) have been validated for all three dimensions. None of those eight wearables were ranked consistently as moderate to strong validity for measuring two or all three dimensions.

DISCUSSION

The purpose of this systematic review was to evaluate the characteristics, validity and quality of free-living validation studies in which at least one dimension of the 24-hour PB construct[1 7] was assessed via wearables and validated against a criterion measure. More specifically, the main purpose was to raise researchers' and consumers' attention to the quality of published validation protocols while aiming to identify and compare specific consistencies/inconsistencies. In summary, we observed a high heterogeneity across the included study protocols. A detailed discussion of each of the points is provided further. Regarding the quality, few studies we

Table 3 Criteria for the risk of bias assessment and the percentage of studies meeting these criteria

Criteria items	Studies meeting criterion (N)			
	Total (N=76)	Biological state (n=12)	Posture/ activity type (n=12)	Intensity (n=52)
Domain 1: patient selection/study design				
1. Was the study conducted in different free-living settings (eg, school or home)?*	43 (67.2%)	NA	7 (58.3%)	36 (69.2%)
2. Did the study take place for at least 2 days?	42 (55.3%)	5 (41.6%)	5 (41.6%)	32 (61.5%)
3. Did the study provide any information about the inclusion/exclusion of the recruiting process?	45 (59.2%)	5 (41.6%)	3 (25%)	37 (71.2%)
4. Did the study include at least a sample of 20 participants?	62 (81.6%)	10 (83.3%)	9 (75%)	43 (82.7%)
Domain 2: index measure				
5. Was the algorithm of the validated outcome reported (ie, formula) or at least further information cited?	34 (44.7%)	9 (75%)	7 (58.3%)	18 (34.6%)
6. Did the participants wear the wearable for at least 8 hours per day?*	46 (60.5%)	NA	4 (33.3%)	32 (61.5%)
Domain 3: criterion measure				
7. Is the selected reference the gold standard?	35 (46.1%)	8 (66.6%)	8 (66.6%)	19 (36.5%)
Domain 4: flow and timing				
8. Did they provide any information about data synchronisation?†	39 (51.3%)	8 (80%)	9 (75%)	20 (54.1%)
9. Were all participants included in the analyses or any exclusion reasons provided?	70 (92.1%)	10 (83.3%)	11 (91.7%)	49 (94.2%)

*Only relevant for n=64 studies.
†Only relevant for n=59 studies.
NA, not applicable.

evaluated were ranked overall with low risk of bias or with some concerns based on selected criteria that align with published core principles, recommendations and expert statements.[16–18] Therefore, more high-quality validation studies with children and adolescents under real-life conditions are needed.

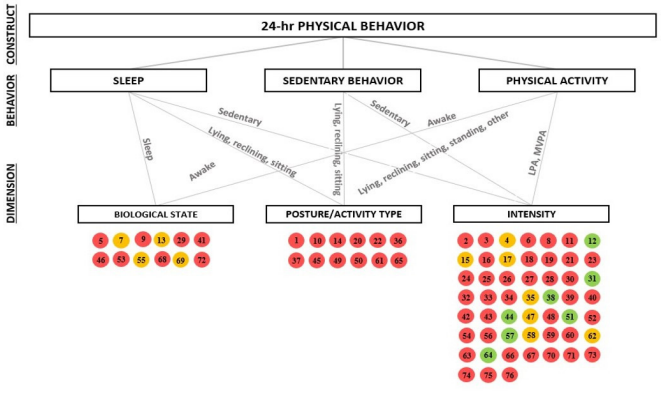

Figure 2 Overall risk of bias classification separated by different dimensions of the 24-hour physical behaviour construct. The number within the circles represents the study number, as listed in the full data extraction (see online supplemental appendices 5 and 8). Studies with a green circle were evaluated as 'low risk'; the orange circle represents 'some concerns'; and red circles represent 'high risk'.

The second purpose of this review was to provide a comprehensive and historical overview on which wearable has been validated for which purpose. In comparison to intensity outcomes such as energy expenditure, the validation of biological state and posture/activity type outcomes was rare. In addition, 42 of 51 different research and consumer-grade wearables were validated for only one aspect of the 24-hour PB construct. We identified only two wearables (ie, ActiGraph GT3X+ and ActiGraph GT1M) that were validated for all three dimensions and six wearables that were validated for at least two dimensions. However, none of those eight wearables were ranked consistently as moderate to strong valid for measuring two or all three dimensions. One of the issues that emerge from the included studies is that while some wearables may be useful for the evaluation of one dimension of the PB construct, we identified no wearable that provides valid results across all three dimensions in children and adolescents.

According to the framework of wearable validation studies,[16] the aim of phase III studies is to validate a device outcome under real-life conditions against appropriate reference measures. Thus, the most central category when

Table 4 Validity of wearables separated by dimensions and age groups

		Early years (infants, preschoolers; <6 years)			School-aged children (6–12 years)			Adolescents (≥13 years)		
Wearable		Bio	Pos/AT	Int	Bio	Pos/AT	Int	Bio	Pos/AT†	Int‡
Uniaxial accelerometer	Actical			↑	↑, ↔					
	ActiGraph AM7164			↔, ↓			↔, ↓; ↓			
	ActiGraph GT1M		↓	↑	↑	↑	↔, ↓			↑
	Actimaker EW4800P									↔
	ActiTrainer						↑			
	Actiwatch activity monitor			↑						
	Actiwatch spectrum			↔				↔		
	Actiwatch-L	↔		↑						
	Actiwatch 64	↓								
	Caltrac						↔, ↓			
	CSA Accelerometer						↑, ↑			
	Motionlogger Sleepwatch				↔					
	New Lifestyle 1000						↑, ↓			
	Omron HJ-151						↑			
	Polar Active Watch						↔			
Biaxial accelerometer	Actiwatch-2	↑, ↓			↔			↔		
	Omron HJ-720IT-E2			↑						
	SenseWear Pro2						↔			
	SenseWear Pro3						↔, ↔			↓
Triaxial accelerometer	3dNX model v2									↑
	3dNX model v3						↑			
	ActiGraph GT3X		↔	↑, ↔						
	ActiGraph GT3X+	↑			↑, ↔	↔, ↓	↑, ↑, ↓	↔	↓	
	ActiGraph GT9X					↔, ↓				
	Actiheart			↔						↔
	ActivPAL		↑, ↑, ↓							
	Axivity AX3					↑				
	Dynaport ADL Monitor					↑				
	Fitbit Charge HR				↔					↔
	Fitbit Flex			↔						
	Fitbit Zip						↔, ↔			
	Garmin Vivofit 1									↑
	GENEActiv					↔, ↓	↔			
	Kenz Lifecorder						↑			
	MovBand Model 2						↔			
	Omron HJA-350IT						↑			
	Polar Loop									↓
	RT3			↔						
	SCA3000		↑							
	SenseWear Mini						↔			
	Sqord						↑			
	Tracmor2						↑			
	TracmorD			↑						
	Tritrac-R3D						↑, ↑, ↓			
	Wristband									↑
	Yamasa EX-200						↑			
	Zamzee						↔			
Accelerometer	Polar A370				↑			↑		

Continued

Table 4 Continued

Wearable		Early years (infants, preschoolers; <6 years)			School-aged children (6–12 years)			Adolescents (≥13 years)		
		Bio	Pos/AT	Int	Bio	Pos/AT	Int	Bio	Pos/AT†	Int‡
Pedometer	Walk4Life MVP						↑, ↔			
	Yamax Digi-Walker SW-200			↑, ↓			↑, ↑, ↔, ↔			
	Yamax Digi-Walker SW-701						↑			

↑ denotes moderate to strong validity; ↔ indicates mixed validity; ↓ indicates poor/weak validity.
bio., biological state; int, intensity; Pos/AT, posture/activity type.

evaluating the overall study quality focused on the selected criterion measure. The evaluation was based on the listed criterion measures for PB assessment by Keadle *et al*.[16] Physiological outcomes such as energy expenditure are recommended for validation against indirect calorimetry or doubly labelled water. Behavioural criterion measures such as step count, or postures are recommended for validation against video recordings.[16] The recommended criterion measure for differentiation between sleep and wake patterns is polysomnography.[111 112] Only 35 of 76 studies used the respective gold standard. Although the relative percentage was higher when the outcome belonged to the biological state or posture/activity type dimension, this might be explained by the fact that only few studies validated either biological state (n=12) or posture/activity type (n=12) outcomes compared with studies that validated intensity outcomes (n=52). The most selected criterion measure was a research-grade device, which may provide information about convergent validity. Although using criterion measures such as video recording can be time-consuming and challenging (eg, low memory capacity or video processing in terms of interpretation), there is no evidence that wearables can serve as a basis for validating other wearables. In other words, if the criterion measure is not completely valid, then a high risk of bias might be present to inform about criterion validity.[15 113 114]

Optimally, study protocols occur over a 24-hour period over multiple days, thus covering a wide range of representative habitual activities.[15 16] First, we evaluated whether data collection was not restricted to one particular setting (eg, school hours), which was met by nearly two-thirds of all reviewed studies. Second, since it is almost not feasible to collect data over several days for criterion measures such as video recording,[15 16] we specified at least 2 days as a low-risk classification. This was met by slightly more than half of the studies. However, we identified a higher number of studies (n=19) that collected data over a short period (≤2 hours). Most of those studies focused on a couple of school hours under free-living conditions while using direct observation without video recordings as a criterion measure. The risk of bias might be present because the setting was restricted to the school environment, thus limiting the ability to capture a wider range of habitual behaviours. Moreover, reactivity reveals a potential error source when collecting data via wearables. Researchers expected that reactivity is a time issue, which means participants may change their behaviour at the beginning of the monitoring period and later return to a more stable pattern.[115 116] Similar effects have been seen in sleep laboratories with polysomnographic monitoring.[117] Since there is some evidence that reactivity of wearing wearables and the first-night effect of polysomnographic data in children and adolescents might be present,[118] we recommend collecting data over at least 2 days.

Ideally, the wearables are validated for a wide range of diverse samples using the same validation protocol (eg, age, sex, ethnicities and health conditions).[16 114] For example, wearables that have been validated for preschoolers have also been validated for adolescents to allow for assessment and comparison across the paediatric age range. The reviewed studies revealed that 53 studies included samples of children between 4 and 13 years of age. In contrast, only 12 studies included newborns and 9 studies included adolescents. Most critically, the majority of the devices were only validated once, which means, for example, that valid results in a sample of preschoolers had not been replicated in infants or adolescents. This finding is in line with a previous review[119] that indicated that current wearable validation studies are limited regarding generalisability in studies with children and adolescents from diverse backgrounds and underrepresented groups. For example, we identified only two studies[57 82] that included samples with restricted health conditions. The practical implication is that a given wearable device might be valid for healthy children and adolescents but not for those with health restrictions.[14] According to the recommended principle, validation study protocols should either include a variety of cohorts within a single study or a series of studies with different participant characteristics.[15 16 120] One solution might be to recruit a larger sample size, which would enable higher intersubject variability, or to conduct a series of validation studies with varying participant characteristics. Most of the reviewed studies included a sample size of at least 20 participants. Optimally, the sample size calculations ensure adequate power for validation purposes.[15 121] Finally, although challenging because of data protection guidelines, we recommend whenever possible reporting information about ethnicity (reported in 21 of 76 studies) and providing detailed information about inclusion/

exclusion criteria regarding the recruitment process as well as for statistical analyses.

To enable comparison between different wearables or wearing locations, researchers may collect data from multiple brands or different wearing positions simultaneously.[16 120] The majority of the reviewed studies did not include multiple wearables and did not capture data from validated devices at different wearing positions. Depending on the primary outcome of interest, the recommendations where to place the wearable may vary. For example, to assess sleep–wake patterns, wrist-worn devices may optimise the recording of small movements that occur at the distal extremities when the individual is supine.[112 122] Notably, compliance issues in studies with younger children might be relevant when selecting the wearing position. Fairclough *et al*[123] reported that wrist placement promotes superior compliance compared with hip placement. If researchers are interested in differentiating between PA and SB, Stevens *et al*[7] indicated that thigh-worn devices might be the most promising position due to the option to wear the device under clothing and accurately assess intensity and posture/activity types. However, only four studies[35 53 61 109] validated posture/activity type outcomes with thigh-worn devices. Future validation studies in children and adolescents are needed with multiple wearables at different wearing positions to increase comparability and to inform end users which device to use and where to place it.[114] In addition, future methods and algorithms might be valuable in terms of extracting and validating different outcomes from a single wearing position.

We evaluated whether the studies reported information concerning data synchronisation, wear time, the algorithm of the validated outcome and data analyses. Overlooking the synchronisation between index and criterion measures may introduce errors and bias the results. Timestamped or pragmatic solutions are recommended, such as participants performing three vertical jumps at the beginning and the end of the measurement.[15] Following practical consideration when applying wearables,[31 124] a high number of studies defined a valid day if ≥8–10 hours of wear time during waking hours were captured. We set the quality criteria of at least ≥8 hours/day, revealing that 46 studies considered wear time criteria for a valid day. Capturing shorter periods may increase the risk of bias since less time is available to assess data in different settings (eg, at home or school). Across all included studies, we identified different outcomes for each dimension. While for the dimension biological state and posture/activity type the outcomes are quite homogenous, the identified intensity outcomes varied from time spent in different intensity to step counts, energy expenditure or metrics such as counts. We included all outcomes that belong to the intensity domain; however, future research endeavours might be interested in further differentiating intensity dimension outcomes in terms of construct validity. A critical aspect from the perspective of transparency is the presentation of algorithms. Only 34 studies reported the formula or cited at least further information on the validated outcome. Interestingly, no information about the algorithms used was provided in studies in which a consumer-grade device had been validated. At this point, researchers often do not have access to the raw data of consumer-grade wearables or the 'black-boxed' algorithms. Moreover, companies can update wearables' firmware or algorithms at any time, hindering comparability.[125] In addition, the pace at which technology is evolving in optimising algorithms far exceeds the pace of published validation research.[20] Open-source methods that are more flexible to use algorithms for different devices are needed.[15 16] A quality criterion concerning the statistical analyses used was not set due to the lack of consistent statistical guidelines for reporting the validity of wearables. The majority of the reviewed studies used traditional statistical tests such as t-tests or ANOVAs. Optimally, researchers integrate different analytical approaches, such as equivalence testing, and include epoch-by-epoch comparisons whenever possible.[16 126]

Future directions

We expect that wearables will be a global surveillance methodology for 24-hour PB assessment.[11 12] Therefore, scientific collaborations such as the ProPASS consortium[127] are fundamental to bundle knowledge and harmonise a currently widely differentiated field of wearable devices. In our review, we identified a high degree of heterogeneity across the study protocols that validated wearables. One reason that may contribute to heterogeneity is the timing of the study realisation. Earlier study protocols may fall short according to quality criteria that have been established over time (eg, opportunity to collect continuous video recordings during activities of daily life). At the latest when the wearables are used as a global surveillance methodology for 24-hour PB assessment, high-quality standards should be maintained (eg, high-quality validation studies). Therefore, in line with previous recommendations,[15 114] we agree that a standardised and transparent validation process should be the primary interest of all stakeholders (ie, manufacturers, scientific institutions and consumers) to assess whether these wearables are useful and perform with low measurement error. The validation framework by Keadle *et al*[16] may serve as such a transparent validation process from device manufacturing to implementation in applied studies. In other words, establishing validity is a process in which multiple pieces of information are needed to confirm validity under different situations (eg, laboratory and free-living) and in different samples (eg, age groups or health conditions), which cannot be accomplished in a single study.[114] Moreover, we expect that the fast development of technical possibilities will influence the future of PB data evaluation and processing via wearables. In particular, supervised learning approaches (eg, machine-learning or deep-learning algorithms) are gaining popularity.[128 129] To date, the uptake of supervised

learning approaches has been slow in health behaviour research and may change in the upcoming years.[12]

Limitations

Some points merit further discussion. First, the evaluation of the study quality is based on self-selected criteria. In particular, we selected the QUADAS-2 tool[104] and added further signalling questions in line with core principles, recommendations and expert statements.[16–18 24] However, since we are not aware of any further quality tools and signalling questions that had been published for wearable validation purposes, our selected criteria can serve as a starting point for future reviews that focus on the study quality of wearable technology under free-living conditions. Second, our included validation studies were published in the range from 1987 to 2021. Given the rapid development of wearable technologies and the increasing availability of different research and consumer-grade devices, quality standards have evolved. Thus, while interpreting the study protocols, the timing of the study realisation should be considered. Moreover, we are aware that most devices were initially not developed for assessing the whole 24-hour PB concept. However, our review can be seen as a comprehensive and historical overview of which wearable had been validated for which purpose and may guide future research endeavours when selecting a wearable for the assessment of the whole 24-hour concept. Third, our review focused on the quality of study protocols. However, we did not account for further important considerations when using wearables such as wear/nonwear time algorithms, cost of the monitor or time of data processing.[114 130] Fourth, our presented narrative data syntheses are based on the author's results/conclusions of the included validation studies. Notably, the overview should be interpreted with caution since study protocols revealed a large heterogeneity in terms of different study protocols (eg, criterion measure, outcomes, sample sizes and duration) or statistical analyses. Fifth, our findings are limited to our search strategy; thus, we may have missed further validation studies. However, we applied backward and forward citation searches through reference lists of the included studies to screen articles that may not have appeared in our search. Sixth, this review was limited to articles published in English.

CONCLUSION

Given the increasing availability of research and consumer-grade wearables, we would like to raise researchers' and consumers' attention to the quality of published validation protocols in children and adolescents. Most reviewed studies did not meet recommended quality principles when validating wearables under real-life conditions. Primarily, validation studies are lacking with gold-standard reference measures such as video recording, polysomnography or doubly labelled water methods. Moreover, most devices had been validated only once and focused predominantly on intensity measure outcomes. Based on reviewed studies, no identified wearable provides valid results for all three dimensions of the 24-hour PB concept in children and adolescents. Since there is a rising interest in the 24-hour PB construct in health research, future researchers will be eager to capture all aspects of PB simultaneously via wearables. It is likely that the next generation of validation studies will consider the validity of more than just one aspect of the 24-hour PB construct during one study protocol or to conduct a series of studies with varying sample characteristics (eg, health status, age and sex). For this purpose, standardised protocols for free-living validation are urgently needed. Standardised protocols embedded in a validation framework may inform and guide all stakeholders (eg, end users, researchers and manufacturers) when (1) selecting wearables for private purposes, (2) applying wearables in health studies or (3) fostering innovation to achieve improved validity.

Author affiliations
[1]Department of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Mannheim, Germany
[3]Department of Sport Pedagogy, University of Bern, Bern, Switzerland
[4]Department of Orthopedics, University Medical Center Rotterdam, Rotterdam, The Netherlands
[5]Department of Rehabilitation Medicine and Physical Therapy, Erasmus University Medical Center, Rotterdam, The Netherlands
[6]Department of Health Science, University of Bern, Bern, Switzerland
[7]Department of Sports and Sports Science, Institute of Sports and Sports Science, Karlsruhe, Germany

Contributors MG and UWE-P contributed to the conception and design of the study. MG, SK, AB and MB contributed to the development of the search strategy. MG, SK, CaN, AB, MB, CIN, ER, A-KD, and IT conducted the systematic review. MG, ER, A-KD, and IT completed the data extraction. All authors assisted with the interpretation. MG, UEWP, JB, CaN, SK, IT and AB were the principal writers of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Marco Giurgiu http://orcid.org/0000-0001-6684-3463

Claudio Nigg http://orcid.org/0000-0002-2897-4689

REFERENCES

1 Rosenberger ME, Fulton JE, Buman MP, *et al*. The 24-hour activity cycle: a new paradigm for physical activity. *Med Sci Sports Exerc* 2019;51:454–64.
2 Tremblay MS, Aubert S, Barnes JD, *et al*. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. *Int J Behav Nutr Phys Act* 2017;14:75.
3 Bull FC, Al-Ansari SS, Biddle S, *et al*. World Health organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med* 2020;54:1451–62.
4 Tremblay MS, Carson V, Chaput J-P, *et al*. Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab* 2016;41:S311–27.
5 Warburton DER, Katzmarzyk PT, Rhodes RE, *et al*. Evidence-Informed physical activity guidelines for Canadian adults. *Can J Public Health* 2007;98 Suppl 2:P16–68.
6 World Health Organization. *Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*. Geneva: World Health Organization, 2019.
7 Stevens ML, Gupta N, Inan Eroglu E, *et al*. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. *BMJ Open Sport Exerc Med* 2020;6:e000874.
8 Keadle SK, Bustamante EE, Buman MP. Physical activity and public health: four decades of progress. *Kinesiol Rev* 2021;10:319–30.
9 Burchartz A, Anedda B, Auerswald T, *et al*. Assessing physical behavior through accelerometry – state of the science, best practices and future directions. *Psychol Sport Exerc* 2020;49:101703.
10 Thompson WR. Worldwide survey of fitness trends for 2022. *ACSMs Health Fit J* 2022;26:11–20.
11 Troiano RP, Stamatakis E, Bull FC. How can global physical activity surveillance adapt to evolving physical activity guidelines? needs, challenges and future directions. *Br J Sports Med* 2020;54:1468–73.
12 Trost SG. Population-Level physical activity surveillance in young people: are accelerometer-based measures ready for prime time? *Int J Behav Nutr Phys Act* 2020;17:28.
13 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
14 Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc* 2012;44:P32–8.
15 Johnston W, Judice PB, Molina García P, *et al*. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med* 2021;55:780–93.
16 Keadle SK, Lyden KA, Strath SJ, *et al*. A framework to evaluate devices that assess physical behavior. *Exerc Sport Sci Rev* 2019;47:206–14.
17 Sperlich B, Holmberg H-C. Wearable, Yes, but able…?: it is time for evidence-based marketing claims! *Br J Sports Med* 2017;51:1240.
18 Mühlen JM, Stang J, Lykke Skovgaard E, *et al*. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE network. *Br J Sports Med* 2021;55:767–79.
19 McCarney R, Warner J, Iliffe S, *et al*. The Hawthorne effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007;7:30.
20 Fuller D, Colwell E, Low J, *et al*. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR Mhealth Uhealth* 2020;8:e18694.
21 Moore CC, McCullough AK, Aguiar EJ, *et al*. Toward harmonized Treadmill-Based validation of Step-Counting wearable technologies: a scoping review. *J Phys Act Health* 2020;17:840–52.
22 Lynch BA, Kaufman TK, Rajjo TI, *et al*. Accuracy of Accelerometers for measuring physical activity and levels of sedentary behavior in children: a systematic review. *J Prim Care Community Health* 2019;10:2150132719874252.
23 Pisanu S, Deledda A, Loviselli A, *et al*. Validity of Accelerometers for the evaluation of energy expenditure in obese and overweight individuals: a systematic review. *J Nutr Metab* 2020;2020:2327017.
24 Sardinha LB, Júdice PB. Usefulness of motion sensors to estimate energy expenditure in children and adults: a narrative review of studies using DLW. *Eur J Clin Nutr* 2017;71:331–9.
25 Ko P-RT, Kientz JA, Choe EK, *et al*. Consumer sleep technologies: a review of the landscape. *J Clin Sleep Med* 2015;11:1455–61.
26 Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Rev Med Devices* 2016;13:497–506.
27 Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Med Rev* 2020;49:101227.
28 Chaput J-P, Willumsen J, Bull F, *et al*. 2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5-17 years: summary of the evidence. *Int J Behav Nutr Phys Act* 2020;17:141.
29 World Health Organization. *Who guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*. World Health Organization, 2019.
30 Baquet G, Stratton G, Van Praagh E, *et al*. Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. *Prev Med* 2007;44:143–7.
31 Migueles JH, Cadenas-Sanchez C, Ekelund U, *et al*. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Med* 2017;47:1821–45.
32 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
33 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
34 Schardt C, Adams MB, Owens T, *et al*. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007;7:16.
35 Alghaeed Z, Reilly JJ, Chastin SFM, *et al*. The influence of minimum sitting period of the ActivPAL™ on the measurement of breaks in sitting in young children. *PLoS One* 2013;8:e71854.
36 Quante M, Kaplan ER, Cailler M, *et al*. Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms. *Nat Sci Sleep* 2018;10:13–20.
37 Smith C, Galland B, Taylor R, *et al*. ActiGraph GT3X+ and Actical wrist and hip worn Accelerometers for sleep and wake indices in young children using an automated algorithm: validation with polysomnography. *Front Psychiatry* 2019;10:958.
38 Alhassan S, Sirard JR, Kurdziel LBF, *et al*. Cross-Validation of two accelerometers for assessment of physical activity and sedentary time in preschool children. *Pediatr Exerc Sci* 2017;29:268–77.
39 Allor KM, Pivarnik JM. Stability and convergent validity of three physical activity assessments. *Med Sci Sports Exerc* 2001;33:671–6.
40 Bäcklund C, Sundelin G, Larsson C. Validity of armband measuring energy expenditure in overweight and obese children. *Med Sci Sports Exerc* 2010;42:1154–61.
41 Barreira TV, Schuna JM, Mire EF, *et al*. Identifying children's nocturnal sleep using 24-h waist accelerometry. *Med Sci Sports Exerc* 2015;47:937–43.
42 Beets MW, Morgan CF, Banda JA, *et al*. Convergent validity of Pedometer and Accelerometer estimates of Moderate-to-Vigorous physical activity of youth. *J Phys Act Health* 2011;8:S295–305.
43 Bélanger Marie-Ève, Bernier A, Paquet J, *et al*. Validating actigraphy as a measure of sleep for preschool children. *J Clin Sleep Med* 2013;9:701–6.
44 De Bock F, Menze J, Becker S, *et al*. Combining accelerometry and HR for assessing preschoolers' physical activity. *Med Sci Sports Exerc* 2010;42:2237–43.
45 Borghese MM, Lin Y, Chaput JP, *et al*. Estimating sleep efficiency in 10- to- 13-year-olds using a waist-worn accelerometer. *Sleep Health* 2018;4:110–5.
46 Busser HJ, Ott J, van Lummel RC, *et al*. Ambulatory monitoring of children's activity. *Med Eng Phys* 1997;19:440–5.
47 Byun W, Kim Y, Brusseau TA. The use of a Fitbit device for assessing physical activity and sedentary behavior in preschoolers. *J Pediatr* 2018;199:35–40.
48 Calabró MA, Stewart JM, Welk GJ. Validation of pattern-recognition monitors in children using doubly labeled water. *Med Sci Sports Exerc* 2013;45:1313–22.
49 Camerota M, Tully KP, Grimes M, *et al*. Assessment of infant sleep: how well do multiple methods compare? *Sleep* 2018;41:zsy146.
50 Carlson JA, Bellettiere J, Kerr J, *et al*. Day-level sedentary pattern estimates derived from hip-worn accelerometer cut-points in 8-12-year-olds: do they reflect postural transitions? *J Sports Sci* 2019;37:1899–909.

51 Carter J, Wilkinson D, Blacker S, *et al*. An investigation of a novel three-dimensional activity monitor to predict free-living energy expenditure. *J Sports Sci* 2008;26:553–61.
52 De Craemer M, De Decker E, Santos-Lozano A, *et al*. Validity of the Omron pedometer and the actigraph step count function in preschoolers. *J Sci Med Sport* 2015;18:289–93.
53 De Decker E, De Craemer M, Santos-Lozano A, *et al*. Validity of the ActivPAL™ and the ActiGraph monitors in preschoolers. *Med Sci Sports Exerc* 2013;45:2002–11.
54 García-Prieto JC, Martinez-Vizcaino V, García-Hermoso A, *et al*. Energy expenditure in playground games in primary school children measured by Accelerometer and heart rate monitors. *Int J Sport Nutr Exerc Metab* 2017;27:467–74.
55 Meltzer LJ, Wong P, Biggs SN, *et al*. Validation of actigraphy in middle childhood. *Sleep* 2016;39:1219–24.
56 Nam Y, Park JW. Physical activity recognition using a single triaxial accelerometer and a barometric sensor for baby and child care in a home environment. *J Ambient Intell Smart Environ* 2013;5:381–402.
57 Voss C, Gardner RF, Dean PH, *et al*. Validity of commercial activity Trackers in children with congenital heart disease. *Can J Cardiol* 2017;33:799–805.
58 Coe D, Pivarnik JM. Validation of the CSA Accelerometer in adolescent boys during Basketball practice. *Pediatr Exerc Sci* 2001;13:373–9.
59 Costello N, Deighton K, Cummins C, *et al*. Isolated & Combined Wearable Technology Underestimate the Total Energy Expenditure of Professional Young Rugby League Players; A Doubly Labelled Water Validation Study. *J Strength Cond Res* 2019. doi:10.1519/JSC.0000000000003434. [Epub ahead of print: 27 Dec 2019].
60 Crouter SE, Oody JF, Bassett DR. Estimating physical activity in youth using an ankle accelerometer. *J Sports Sci* 2018;36:2265–71.
61 Davies G, Reilly JJ, McGowan AJ, *et al*. Validity, practical utility, and reliability of the activPAL™ in preschool children. *Med Sci Sports Exerc* 2012;44:761–8.
62 Djafarian K, Speakman JR, Stewart J, *et al*. Comparison of activity levels measured by a wrist worn accelerometer and direct observation in young children. *Open J Pediatr* 2013;03:422–7.
63 Duncan S, White K, Sa'ulilo L, *et al*. Convergent validity of a piezoelectric pedometer and an omnidirectional accelerometer for measuring children's physical activity. *Pediatr Exerc Sci* 2011;23:399–410.
64 Finn KJ, Specker B. Comparison of Actiwatch activity monitor and children's activity rating scale in children. *Med Sci Sports Exerc* 2000;32:1794–7.
65 Godino JG, Wing D, de Zambotti M, *et al*. Performance of a commercial multi-sensor wearable (Fitbit charge HR) in measuring physical activity and sleep in healthy children. *PLoS One* 2020;15:e0237719.
66 Grydeland M, Hansen BH, Ried-Larsen M, *et al*. Comparison of three generations of ActiGraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports Sci Med Rehabil* 2014;6:26.
67 Hallal PC, Reichert FF, Clark VL, *et al*. Energy expenditure compared to physical activity measured by accelerometry and self-report in adolescents: a validation study. *PLoS One* 2013;8:e77036.
68 Hands B, Larkin D. Physical activity measurement methods for young children: a comparative study. *Meas Phys Educ Exerc Sci* 2006;10:203–14.
69 Hart TL, Brusseau T, Kulinna PH, *et al*. Evaluation of low-cost, objective instruments for assessing physical activity in 10-11-year-old children. *Res Q Exerc Sport* 2011;82:600–9.
70 Hislop JF, Bulley C, Mercer TH, *et al*. Comparison of epoch and uniaxial versus triaxial accelerometers in the measurement of physical activity in preschool children: a validation study. *Pediatr Exerc Sci* 2012;24:450–60.
71 Hurter L, Fairclough SJ, Knowles ZR, *et al*. Establishing RAW acceleration thresholds to classify sedentary and stationary behaviour in children. *Children* 2018;5:172.
72 Hurter L, Rowlands AV, Fairclough SJ, *et al*. Validating the sedentary sphere method in children: does wrist or accelerometer brand matter? *J Sports Sci* 2019;37:1910–8.
73 Kim Y, Lochbaum M. Comparison of Polar Active Watch and Waist- and Wrist-Worn ActiGraph Accelerometers for Measuring Children's Physical Activity Levels during Unstructured Afterschool Programs. *Int J Environ Res Public Health* 2018;15:2268.
74 Klesges LM, Klesges RC. The assessment of children's physical activity: a comparison of methods. *Med Sci Sports Exerc* 1987;19:511–7.
75 Krishnaveni GV, Mills IC, Veena SR, *et al*. Accelerometers for measuring physical activity behavior in Indian children. *Indian Pediatr* 2009;46:1055–62.
76 Ramírez-Marrero FA, Smith BA, Sherman WM, *et al*. Comparison of methods to estimate physical activity and energy expenditure in African American children. *Int J Sports Med* 2005;26:363–71.
77 Ishikawa-Takata K, Kaneko K, Koizumi K, *et al*. Comparison of physical activity energy expenditure in Japanese adolescents assessed by EW4800P triaxial accelerometry and the doubly labelled water method. *Br J Nutr* 2013;110:1347–55.
78 Reilly JJ, Kelly LA, Montgomery C, *et al*. Validation of Actigraph accelerometer estimates of total energy expenditure in young children. *Int J Pediatr Obes* 2006;1:161–7.
79 Welk GJ, Corbin CB, Kampert JB. The validity of the Tritrac-R3D activity monitor for the assessment of physical activity: II. temporal relationships among objective assessments. *Res Q Exerc Sport* 1998;69:395–9.
80 Treuth MS, Sherwood NE, Butte NF, *et al*. Validity and reliability of activity measures in African-American girls for gems. *Med Sci Sports Exerc* 2003;35:532–9.
81 Tanaka C, Hikihara Y, Inoue S, *et al*. The choice of pedometer impacts on daily step counts in primary school children under free-living conditions. *Int J Environ Res Public Health* 2019;16:4375.
82 Sitnick SL, Goodlin-Jones BL, Anders TF. The use of actigraphy to study sleep disorders in preschoolers: some concerns about detection of nighttime awakenings. *Sleep* 2008;31:395–401.
83 Sirard JR, Masteller B, Freedson PS, *et al*. Youth oriented activity Trackers: comprehensive Laboratory- and Field-Based validation. *J Med Internet Res* 2017;19:e250.
84 Silva DR, Minderico CS, Júdice PB, *et al*. Agreement between GT3X Accelerometer and ActivPAL Inclinometer for estimating and detecting changes in different contexts of sedentary time among adolescents. *J Phys Act Health* 2019;16:780–4.
85 Sijtsma A, Schierbeek H, Goris AHC, *et al*. Validation of the TracmorD triaxial accelerometer to assess physical activity in preschool children. *Obesity* 2013;21:1877–83.
86 Santos DA, Silva AM, Matias CN, *et al*. Validity of a combined heart rate and motion sensor for the measurement of free-living energy expenditure in very active individuals. *J Sci Med Sport* 2014;17:387–93.
87 Ruch N, Rumo M, Mäder U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. *Eur J Appl Physiol* 2011;111:1917–27.
88 Rowlands AV, Eston RG. Comparison of accelerometer and pedometer measures of physical activity in boys and girls, ages 8-10 years. *Res Q Exerc Sport* 2005;76:251–7.
89 Rodriguez G, Béghin L, Michaud L, *et al*. Comparison of the TriTrac-R3D accelerometer and a self-report activity diary with heart-rate monitoring for the assessment of energy expenditure in children. *Br J Nutr* 2002;87:623–31.
90 Crouter SE, Flynn JI, Bassett DR. Estimating physical activity in youth using a wrist accelerometer. *Med Sci Sports Exerc* 2015;47:944–51.
91 Mooses K, Oja M, Reisberg S, *et al*. Validating Fitbit ZIP for monitoring physical activity of children in school: a cross-sectional study. *BMC Public Health* 2018;18:858.
92 Ettienne R, Nigg CR, Li F, *et al*. Validation of the Actical Accelerometer in multiethnic preschoolers: the children's healthy living (Chl) program. *Hawaii J Med Public Health* 2016;75:95.
93 Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. *Med Sci Sports Exerc* 1994;26:369–75.
94 Kinder JR, Lee KA, Thompson H, *et al*. Validation of a hip-worn accelerometer in measuring sleep time in children. *J Pediatr Nurs* 2012;27:127–33.
95 Kwon S, Zavos P, Nickele K, *et al*. Hip and wrist-worn accelerometer data analysis for toddler activities. *Int J Environ Res Public Health* 2019;16:2598.
96 Migueles JH, Delisle Nyström C, Henriksson P, *et al*. Accelerometer data processing and energy expenditure estimation in preschoolers. *Med Sci Sports Exerc* 2019;51:590–8.
97 Ojiambo R, Konstabel K, Veidebaum T, *et al*. Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: the IDEFICS validation study. *J Appl Physiol* 2012;113:1530–6.
98 Schneider M, Chau L. Validation of the Fitbit ZIP for monitoring physical activity among free-living adolescents. *BMC Res Notes* 2016;9:448.
99 Šimůnek A, Dygrýn J, Jakubec L, *et al*. Validity of Garmin Vívofit 1 and Garmin Vívofit 3 for school-based physical activity monitoring. *Pediatr Exerc Sci* 2019;31:130–6.

100 Unno M, Morisaki T, Kinoshita M, *et al*. Validation of actigraphy in hospitalised newborn infants using video polysomnography. *J Sleep Res* 2022;31:e13437.
101 Gao Z, Lee AM, Solmon MA. Validating Pedometer-based physical activity time against Accelerometer in middle school physical education. *ICHPER-SD Journal of Research* 2010;5:20–5.
102 Pulakka A, Cheung YB, Ashorn U, *et al*. Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatr* 2013;102:1192–8.
103 Pesonen A-K, Kuula L. The validity of a new Consumer-Targeted wrist device in sleep measurement: an overnight comparison against polysomnography in children and adolescents. *J Clin Sleep Med* 2018;14:585–91.
104 Oliver M, Schofield GM, Kolt GS, *et al*. Pedometer accuracy in physical activity assessment of preschool children. *J Sci Med Sport* 2007;10:303–10.
105 Krishnaveni GV, Veena SR, Kuriyan R, *et al*. Relationship between physical activity measured using accelerometers and energy expenditure measured using doubly labelled water in Indian children. *Eur J Clin Nutr* 2009;63:1313–9.
106 Hoos MB, Plasqui G, Gerver W-JM, *et al*. Physical activity level measured by doubly labeled water and accelerometry in children. *Eur J Appl Physiol* 2003;89:624–6.
107 Rowlands AV, Rennie K, Kozarski R, *et al*. Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc* 2014;46:2308–16.
108 Yang X, Jago R, Zhang Q, *et al*. Validity and reliability of the Wristband activity monitor in free-living children aged 10-17 years. *Biomed Environ Sci* 2019;32:812–22.
109 Narayanan A, Stewart T, Mackay L. A Dual-Accelerometer system for detecting human movement in a free-living environment. *Med Sci Sports Exerc* 2020;52:252–8.
110 Van Hoye A, Nicaise V, Sarrazin P. Self-Reported and objective physical activity measurement by active youth. *Sci Sports* 2014;29:78–87.
111 Grandner MA, Rosenberger ME. Actigraphic sleep tracking and wearables: historical context, scientific applications and guidelines, limitations, and considerations for commercial sleep devices. In: *Sleep and health*. Elsevier, 2019: 147–57.
112 Ancoli-Israel S, Martin JL, Blackwell T, *et al*. The SBSM guide to actigraphy monitoring: clinical and research applications. *Behav Sleep Med* 2015;13 Suppl 1:S4–38.
113 Umemneku Chikere CM, Wilson K, Graziadio S, *et al*. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. *PLoS One* 2019;14:e0223832.
114 Welk GJ, Bai Y, Lee J-M, *et al*. Standardizing analytic methods and reporting in activity monitor validation studies. *Med Sci Sports Exerc* 2019;51:1767–80.
115 Clemes SA, Matchett N, Wane SL. Reactivity: an issue for short-term pedometer studies? *Br J Sports Med* 2008;42:68–70.
116 Rowe DA, Mahar MT, Raedeke TD, *et al*. Measuring physical activity in children with Pedometers: reliability, reactivity, and replacement of missing data. *Pediatr Exerc Sci* 2004;16:343–54.
117 Driller MW, Dunican IC. No familiarization or 'first-night effect' evident when monitoring sleep using wrist actigraphy. *J Sleep Res* 2020:e13246.
118 Verhulst SL, Schrauwen N, De Backer WA, *et al*. First night effect for polysomnographic data in children and adolescents with suspected sleep disordered breathing. *Arch Dis Child* 2006;91:233–7.
119 Phillips SM, Summerbell C, Hobbs M, *et al*. A systematic review of the validity, reliability, and feasibility of measurement tools used to assess the physical activity and sedentary behaviour of pre-school aged children. *Int J Behav Nutr Phys Act* 2021;18:1–28.
120 Freedson P, Bowles HR, Troiano R, *et al*. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc* 2012;44:S1–4.
121 Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc* 2012;44:S61–7.
122 Quante M, Kaplan ER, Rueschman M, *et al*. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. *Sleep Health* 2015;1:275–84.
123 Fairclough SJ, Noonan R, Rowlands AV, *et al*. Wear compliance and activity in children wearing Wrist- and Hip-Mounted Accelerometers. *Med Sci Sports Exerc* 2016;48:245–53.
124 Troiano RP, Berrigan D, Dodd KW, *et al*. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;40:181–8.
125 Feehan LM, Geldman J, Sayre EC, *et al*. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth Uhealth* 2018;6:e10527.
126 Dixon PM, Saint-Maurice PF, Kim Y, *et al*. A primer on the use of equivalence testing for evaluating measurement agreement. *Med Sci Sports Exerc* 2018;50:837–45.
127 Stamatakis E, Koster A, Hamer M, *et al*. Emerging Collaborative research platforms for the next generation of physical activity, sleep and exercise medicine guidelines: the prospective physical activity, sitting, and sleep Consortium (ProPASS). *Br J Sports Med* 2020;54:435–7.
128 Dunn J, Kidzinski L, Runge R, *et al*. Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nat Med* 2021;27:1105–12.
129 Farrahi V, Niemelä M, Kangas M, *et al*. Calibration and validation of accelerometer-based activity monitors: a systematic review of machine-learning approaches. *Gait Posture* 2019;68:285–99.
130 Choi L, Liu Z, Matthews CE, *et al*. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc* 2011;43:357–64.