Journal Pre-proof

The wearable landscape: Issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment

Marco Giurgiu , Birte von Haaren-Mack , Janis Fiedler , Simon Woll , Alexander Burchartz , Simon Kolb , Sascha Ketelhut , Claudia Kubica , Carina Nigg , Irina Timm , Maximiliane Thron , Steffen Schmidt , Kathrin Wunsch , Gerhard Müller , Claudio R. Nigg , Alexander Woll , Markus Reichert , Ulrich Ebner-Priemer , Johannes BJ Bussmann

PII: S2095-2546(24)00164-9
DOI: https://doi.org/10.1016/j.jshs.2024.101006
Reference: JSHS 101006

To appear in: *Journal of Sport and Health Science*

Received date: 20 February 2024
Revised date: 24 April 2024
Accepted date: 4 July 2024

Please cite this article as: Marco Giurgiu , Birte von Haaren-Mack , Janis Fiedler , Simon Woll , Alexander Burchartz , Simon Kolb , Sascha Ketelhut , Claudia Kubica , Carina Nigg , Irina Timm , Maximiliane Thron , Steffen Schmidt , Kathrin Wunsch , Gerhard Müller , Claudio R. Nigg , Alexander Woll , Markus Reichert , Ulrich Ebner-Priemer , Johannes BJ Bussmann , The wearable landscape: Issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment, *Journal of Sport and Health Science* (2024), doi: https://doi.org/10.1016/j.jshs.2024.101006

JSHS_101006

Dochead: Opinion

The wearable landscape: Issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment

Marco Giurgiu [a,*] , Birte von Haaren-Mack [a], Janis Fiedler [a], Simon Woll [a], Alexander Burchartz [a], Simon Kolb [a], Sascha Ketelhut [b], Claudia Kubica [b], Carina Nigg [c], Irina Timm [a], Maximiliane Thron [a], Steffen Schmidt [a], Kathrin Wunsch [a], Gerhard Müller [a,d], Claudio R. Nigg [b], Alexander Woll [a], Markus Reichert [e,#], Ulrich Ebner-Priemer [a,#] Johannes BJ Bussmann [f,#]

[a] Institute of Sports and Sports Science, Karlsruhe Institute of Technology (KIT), Karlsruhe 76131, Germany

[b] Institute of Social and Preventive Medicine, University of Bern, Bern 3012, Switzerland

[c] Department of Health Science, Institute of Sport Science, University of Bern, Bern 3012, Switzerland

[d] Allgemeine Ortskrankenkasse AOK Baden-Wuerttemberg, Stuttgart 70191, Germany

[e] Department of eHealth and Sports Analytics, Faculty of Sport Science, Ruhr University Bochum (RUB), Bochum 44801, Germany

[f] Department of Rehabilitation Medicine, Erasmus University Medical Center, Rotterdam 3015, The Netherlands

Running title:
Validation of wearables for the assessment of 24-h physical behavior

*Corresponding author.
E-mail: marco.giurgiu@kit.edu (Giurgiu. M)
[#]Joint last authors

Received 20 February 2024 ; revised 24 April 2024; accepted 4 July 2023

The ongoing revolution in information technology is reshaping human life. In the realm of health behavior, wearable technology emerges as a leading digital solution, capturing physical behaviors (i.e., physical activity, sedentary habits, sleep patterns) within the 24-h cycle of daily life. Wearables are applied in research, clinical practice, and as lifestyle devices; most obvious, they promise to be a key element for increasing human physical activity, one of the biggest health challenges nowadays. However, despite the high expectations associated with wearable technology, fundamental aspects remain surprisingly neglected. Here, the lack of methodologically sound validation studies for wearables entering the market appears particularly critical. In our recent and comprehensive review of 967 validation studies, [1–4] the overall low study quality was evident and alarming. In essence, standard regulatory processes are missing, although the scientific community is strongly advocating for improvements in the validation and trustworthiness of digital health products. Hence, we call for validity of wearables to be systematically tested to pave the way towards expedient digital health solutions and, in terms of reproducibility, to provide transparent information about the devices used (i.e., all data processing steps, analytical approaches, updates of algorithms). Toward this end, our viewpoint compiles challenges and suggests key elements for enhancing the quality of validation protocols (i.e., wearing position, criterion measurement, validated outcomes, sample size, statistical analyses) as well as issues pertinent to improving the validation process (i.e., replication of studies, access to raw data, the release of a new version). Moreover, to catalyze this comprehensive validation process, we are launching and introducing the project Wearable Landscape (www.wearable-landscape.info); the open science initiative not only compiles validation protocols but also facilitates collaboration efforts for sharing resources, equipment, set-up of multi-location studies, as well as joint data analysis and pre-processing.

1. Relevance of validity when using wearables to measure 24-h physical behavior

"People sleep less, sit more, walk less frequently, drive more regularly, and do less physical activity than they used to".[5] This quote describes the lifestyle of young people in the 21st century, emphasizing central risk factors that contribute to poor health, low quality of life, morbidity, and premature mortality.[6] Empirical evidence shows that 24-h physical behavior—which encompasses all movement and posture-related physical behaviors, including physical activity, sedentary, and sleep behavior[7]—is strongly associated with mental and physical health conditions across all age groups throughout the lifespan.[8] Digital health solutions such as wearable technologies are a promising application for increasing human physical activity, one of the biggest current health challenges. Thus, to investigate associations between physical behavior and health, researchers have started gathering comprehensive data on postural and movement patterns in everyday life via wearable technology. The purpose of their application is manifold and could include use as an observational tool in surveillance studies,[9] as a motivational tool in interventions,[10] as a way to better understand the underlying mechanisms of health, treatment, and recovery (i.e., understanding the role of physical behavior), and as a diagnostic tool in clinical settings.[11] Moreover, as one of the first digital health solutions for capturing physical behavior, wearables have become widely used lifestyle devices, with a growing consumer market and millions of users around the globe.[12] The evolution of wearables as an integral part of research and public use can be seen as a blueprint for other digital health solutions.

The validity of wearables is crucial for developing wearable-based public health guidelines[13] and effective prescription of 24-h physical behavior. Validity, in this context, denotes the degree to which a wearable precisely measures the intended values—a critical criterion for evaluating the quality of a wearable.[14] If the devices do not measure precisely, the user information will be incorrect and thus may have detrimental instead of beneficial effects on

health. For example, a user aims to move a recommended amount of 9000 steps per day[15] while wearing a wrist-worn wearable to monitor daily steps. If users are relying on displayed feedback and reach their goals, they may reduce their activity levels for the rest of the day. However, as highlighted by recent validation studies,[16–18] step estimates are often imprecise; therefore, it's plausible that 9000 steps could be an overestimation when an individual has only achieved 6000 steps. In a case such as this, the feedback could potentially harm the individual—in a meta-analysis examining steps and mortality, a difference in 3000 steps/day was associated with a 30% increase in mortality risk.[19] However, to date, market access for wearables through certification or authorization relies on regulatory commissions (e.g., European CE (European conformity)-marking or U.S. Food and Drug Administration), which focus primarily on aspects such as the privacy or security of the processed data.[20] While these regulatory commissions advocate for enhanced interoperability through the ongoing development and validation of digital tools,[21] they do not enforce regulations explicitly targeting the technical validation of digital tools such as wearables. For example, the Digital Medicine Society is seeking to develop clinical-quality resources on a tech timeline and to deliver these actionable resources to the field via open-source channels and educational programs.[22] As a result, a wearable landscape with no quality criteria regarding the refinement and optimization of a device across different phases of validation emerged.

Accordingly, in the realm of wearable technology, we urgently call for the regulation of high-quality validation protocols that are embedded in a phase-based validation framework, as presented by Keadle and colleagues.[23] We recently conducted a series of systematic reviews of 967 validation studies addressing the quality of validation protocols for the assessment of 24-h physical behavior via wearables.[1–4] We identified 391 different wearables from 166 different brands (Fig. 1). Of note, the number of validation studies does not provide any information about the validity of the devices itself. This wearable landscape viewpoint aimed

to illustrate the quality of validation studies, extract challenges, and offer recommendations for improving validation processes and protocols, as well as to discuss future solutions.

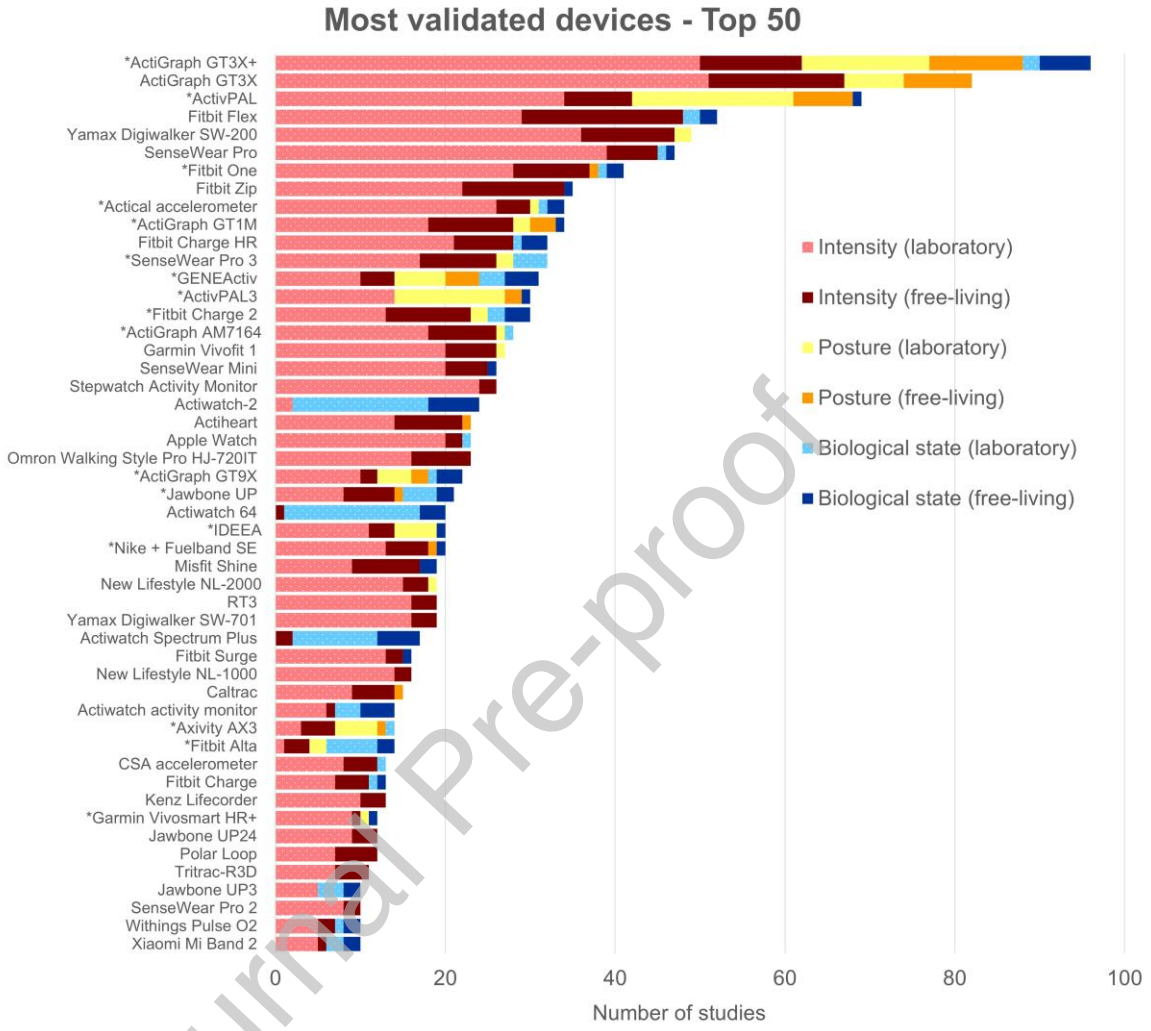

Fig. 1. The 50 most validated devices categorized by the dimensions of the 24-h physical behavior (i.e., red: intensity; yellow: posture/activity types; blue: biological state) and contextual setting (i.e., bright: laboratory; dark: free-living). The asterisk indicates that the devices validated parameters from all dimensions of the 24-h physical behavior construct.

2. Quality of validation studies for the assessment of 24-h physical behavior via wearables

To unveil the validity of the entire landscape of wearables, we summarized the 4 systematic reviews we recently published.[1–4] In these reviews, studies were primarily categorized by

target group (children/adolescents *vs*. adults), setting of the validation study (laboratory, i.e., protocols with structured and semi-structured assessment *vs*. free-living, i.e., real-world conditions where participants can perform their natural daily behaviors), and 24-h physical behavior construct (i.e., intensity dimension, such as energy expenditure or steps; posture/activity types dimension, such as sitting, standing, or walking; and biological state dimension, such as the state of being awake or asleep).[7] In a comprehensive effort, we reviewed the quality of 967 studies [1–4] that validated the results of at least 1 dimension of the 24-h physical behavior construct. Reflecting the pace of technological advancements of wearable technology, most validation studies were published after 2011 (77.7%). Across all studies, the main target population was adults 18–64 years of age (68.1%), followed by older adults (≥65 years; 12.2 %), and children 8–13 years of age (10.9 %). Review of studies by age group showed a paucity of studies in adolescents compared to other age groups., which is a gap in the validation literature that should be addressed in future research. (For more details see Table 1 and published reviews [1–4]).

The overall quality of a study was evaluated using a modified version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2).[24] In particular, the tool comprises 4 domains (i.e., patient selection, index measure, criterion measure, and flow/timing). Following the QUADAS-2 guidelines, we selected a set of signaling 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.[23–26] Notably, since we are not aware of any further quality tools and signaling questions that had been published for wearable validation purposes, our selected criteria can serve as a starting point for future reviews. The risk of bias for each study was categorized into low risk (5.3%), some concerns (18.0%), or high risk (76.7%). One-third (31.5%) of all studies were rated as high risk, because the selected wearables were not validated against a criterion measure aligned

with recommended gold standards,[23] such as indirect calorimetry (28.7%), video recordings (20.1%), or polysomnography (13.7%). Further characteristics of the existing studies are illustrated in Fig. 2. Most studies were conducted with a focus on the intensity dimension, with some low-risk studies for both settings (laboratory and free-living) and target groups (children/adolescents and adults). The dimension of biological state had a lower overall number of studies, and no studies in the laboratory setting were classified as low risk. In contrast, studies focusing on posture and activity types in children were only conducted in laboratory settings, with high-risk studies dominating; only a limited number of low-risk studies and some concerning studies were available.

The status quo of the quality of validation studies is considered critical: There are no existing regulatory mechanisms or guidelines for the technical validation of a wearable to assess either the whole 24-h physical behavior construct or for single dimensions. Furthermore, it is concerning that many previous validation studies are considered to be low quality due to bias.

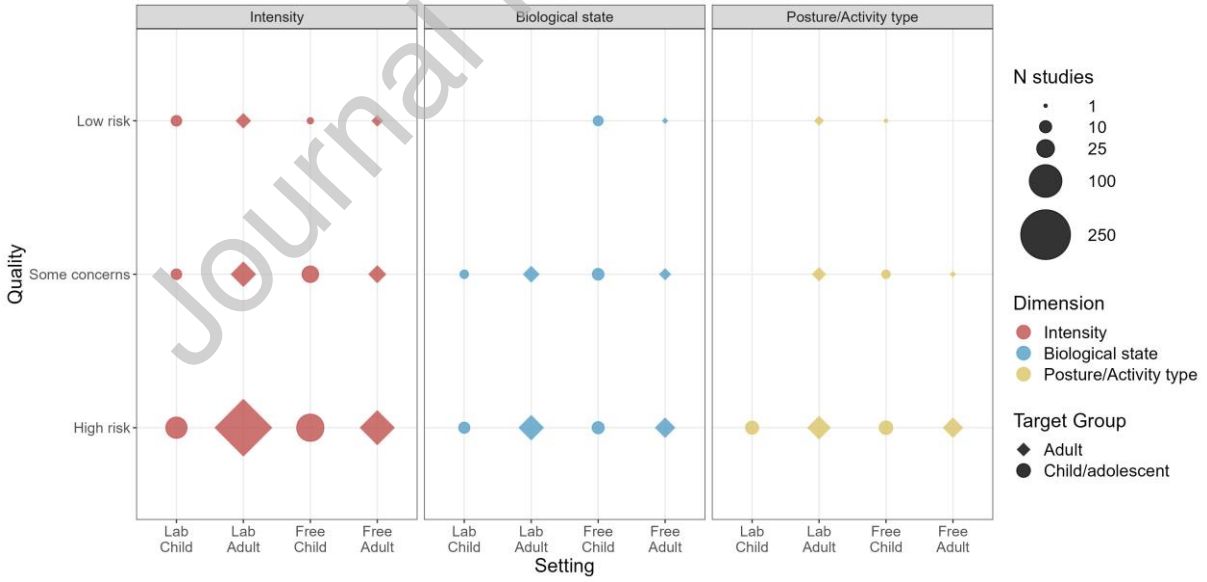

Fig. 2. Quality of the validation protocols categorized by 3 dimensions, settings, populations, and target groups. The number of studies is indicated by the size of the circle or square. Free = free-living; Lab = laboratory

Table 1. Statistical summary of 967 published validation studies, categorized for each dimension of the 24-h physical behavior construct.

	Total	Intensity		Posture/activity type		Biological state	
		Laboratory	Free-living	Laboratory	Free-living	Laboratory	Free-living
Target group							
Pre-school children (≤4 years)	17		6	1	2	4	4
School children (5–12 years)	138	63	38	10	8	12	7
Adolescents (13–17 years)	36	16	7	2		9	2
Adults (≥18 years)	808	420	154	70	36	80	48
Study location							
Africa	4	1	2	---	---	1	---
Asia	67	33	15	1	3	9	6
Europe	382	175	89	54	29	19	16
North America	456	252	86	19	11	58	30
Australia/Oceania	105	44	12	14	5	22	8
South America	6	5	1	---	---	---	---
Number of participants							
≤19	278	110	50	32	26	18	42
20–50	664	331	114	45	20	59	95
≥51	181	69	42	11	2	28	29
Sex assigned at birth (female, %)							
0–25	103	42	24	11	7	13	6
26–74	737	370	147	61	30	80	49
75–100	122	62	25	13	8	9	5
Criterion measure [a]							
Doubly labeled water	57	---	57	---	---	---	---
Indirect calorimetry	283	267	8	8	---	---	---
Observation (direct)	173	128	14	26	4	1	---
Observation (video)	210	129	6	53	19	1	2
Polysomnography	132	---	---	---	---	101	31
Questionnaire/diary	21	1	7	---	4	1	8
Wearable	198	45	114	4	19	---	16
Outcome [b]							
Sleep time/metrics	166	---	---	---	---	106	60
Different postures/types	125	---	1	88	36	---	---
Intensity time (SB)	48	10	38	---	---	---	---
Intensity time (LPA)	34	10	24	---	---	---	---
Intensity time (MVPA)	70	18	52	---	---	---	---
Energy expenditure	348	255	93	---	---	---	---

Steps	359	268	91	---	---	---	---
Counts	64	47	16	---	1	---	---

[a] Criterion measures that were not listed but identified through the review process: Heart telemetry (*n* = 1); 3-dimensional gait analysis (*n* = 1); Direct calorimetry (*n* = 1); Observation (images *n* = 2); Compendium (*n* = 4); EEG or Zmachine (*n* = 4).

[b] Outcomes that were not listed but identified through the review process: Sit-to-Stand transitions (*n* = 5); Time spent in physical activity (*n* = 6); Time spent in walking/active (*n* = 3).

Abbreviations: LPA = low intensity physical activity ; MVPA = moderate to vigorous intensity physical activity; SB = sedentary behavior.

Ideally a wearable undergoes a structured validation process, from manufacturing to its implementation in applied studies. A phase-based validation framework[23] serves as a starting point with both fixed and semi-structured protocols in the laboratory (Phase II) as well as in naturalistic conditions (Phase III), where participants can engage in their natural everyday physical behaviors. The framework establishes the foundation for a validation process in which high-quality protocols should be embedded. Toward this end, we selected issues warranting further discussion to enhance the quality of validation protocols (i.e., wearing position, criterion measurement, validated outcomes, sample size, statistical analyses) and issues pertinent to improving the validation process (i.e., replication of studies, access to raw data, the release of a new version). We would like to remark that our suggestions are based on published reviews, expert statements, and core principles,[1–4,23,25,26] but should not be viewed as direct evidence-based recommendations.

3. Issues that researchers should consider to improve the quality of a validation protocol

3.1 The importance of the criterion measure.

In 68.5% of all studies, wearables were validated against a criterion measure aligning with recommended gold standards,[23] such as indirect calorimetry, video recordings, or polysomnography. We identified substantial differences between the percentages of studies including gold-standard criterion measurements in the laboratory (77.3% protocols) and under free-living conditions (40.2% protocols). For example, a considerable number of studies (19.4%) compared wearables with each other, evaluating concurrent validity rather than assessing criterion validity against a ground truth. This is particularly problematic if the device used for comparison has not been validated for the specific target group, setting, and/or

outcome of interest. Therefore, to promote high-quality protocols, recommended gold standard measures must be used.[23] However, researchers should be aware that the integration of gold standard methods is sometimes accompanied by massive efforts (e.g., observation in the laboratory or counting video-recorded steps) and high costs (e.g., doubly labeled water method) and, especially in free-living protocols, is often difficult to apply in larger samples. Importantly, the complex nature of human movement[27] implies that researchers in the field have not yet agreed on a universal gold standard. This neglect is plausibly attributable to the challenge of capturing the physics of physical behavior *vs.* the physiology of physical behavior (i.e., the behavior *vs.* the physiological response). In other words, the physical behavior can be captured partly by using, for example, accelerometers or motion capture systems, whereas the physiology can be assessed partly by using measures such as maximum oxygen uptake or lactate.

There is a scarcity of studies that include multiple aspects of physical behavior and criterion measures. The majority of protocols were designed to validate an intensity parameter, such as energy expenditure or steps (71.7%). This was followed by validations for biological state outcomes (16.3%) and posture/activity type parameters (13.6%). Across all devices analyzed, we identified 23 wearables that validated parameters from each domain (see asterisk in Fig. 1). The 24-h physical behavior construct has started gaining more attention in the community.[28] During the review process, we did not account for the detection of wear and non-wear time as a critical characteristic of 24-h physical behavior analyses. Here, compositional data analyses[29] offer valuable insights into the association between compositions of 24-h physical behavior and health parameters. Consequently, researchers might be interested in using a single device capable of collecting all domains simultaneously rather than employing a multisensor system from different brands. However, researchers may be interested in using a multisensor system—for example, combining different sensors such as acceleration, gyroscope, heart rate, electrocardiograph, electrodermal activity, and global

positioning system. If researchers are interested in capturing the 24-h physical behavior construct by using a single device, we encourage research endeavors to expand the protocols to validate different outcomes (e.g., postures and steps) within a single protocol, thus pooling resources.[23]

3.2 Planning the sample size.

Most studies recruited sample sizes ranging from 20 to 50 participants (59.0%). According to recommendations from the INTERLIVE network,[30] the sample size should be determined by an *a priori* power estimation, or, if previous data are not available, by recruiting at least 45 participants. Since the objective of validation studies is usually to yield generalizable results, larger and more heterogeneous samples are required. Most importantly, the sample of the validation study must be representative for the target population of interest. We encourage researchers to conduct an *a priori* power analysis and to recruit heterogeneous samples with respect to factors such as age, sex, race, body mass, and atypical gait patterns, thereby facilitating potential subgroup analyses.

3.3 Economizing resources and integrating more wearing positions and brands.

In 57.2% of all validation studies, researchers validated a single device against a criterion measure, with wrist and hip/waist being the most common wearing positions (Fig. 3). According to recommendations, wrist placement is preferred for detecting wake and sleep states due to small movements at the distal extremities in a supine position.[31] Moreover, wrist-worn devices are prone to enhance compliance due to the possibility of receiving real-time feedback. The hip placement is effective for proximity to the center of mass, capturing gross muscle movements like walking or running, and detecting body acceleration and deceleration.[32] Further, thigh placement is currently the most promising position for accurately capturing all posture/activity types.[7] However, with advancements in artificial intelligence, it is increasingly possible to derive sleep and physical activity phenotypes from

sensor data. For example, a recent study across 700,000 person-days of wearable data revealed the potential of self-supervised learning approaches for human activity recognition even in data with limited labeling or where good sampling coverage is hard to achieve.[33] Notably, researchers identified large differences while comparing activity parameters from different sensor positions.[16] Considering these factors and the economic use of resources, researchers might consider adding 3 or 4 brands of wearables or different wearing positions of the same brand to enable comparisons between different wearables and positions.[23]

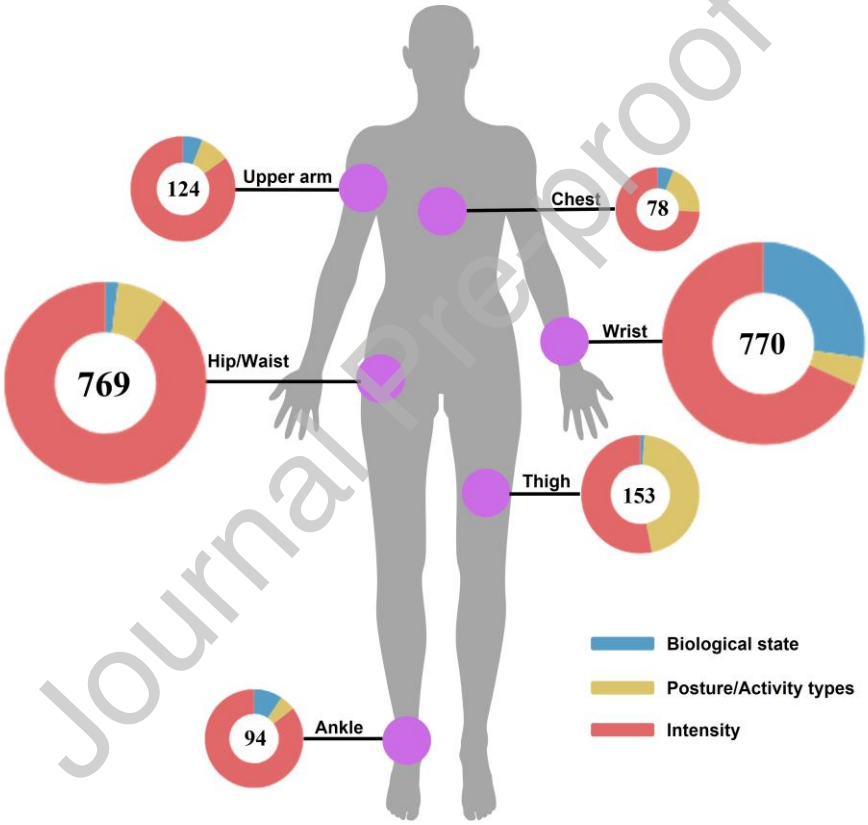

Fig. 3. The 6 most validated wearing positions and the number of devices tested at the respective wearing position, as well as the distribution of the number of studies for each domain of the 24-h physical behavior construct (i.e., red: intensity; yellow: posture/activity types; blue: biological state).

3.3 The choice of appropriate statistical analyses.

Most studies analyzed the data on an aggregated person level by using Bland-Altman analysis, correlations, *t* tests, and analyses of variance. Further approaches, such as equivalence testing, are recommended.[23,34] For example, Dixon et al.[35] provided principles of equivalence testing for various scenarios and indicated that an equivalence zone of ±15% is applicable for comparing criterion measures with wearable measures. In addition to these analyses, higher data granulation (e.g., epoch-by-epoch comparisons) focusing on higher data resolution (e.g., seconds) can provide valuable insights into within-device differences from moment to moment. A device might be valid on an aggregated day-level scale but at the same time reveal a lack of validity on min-by-min evaluations. In line with recently published recommendations,[36] we emphasize that researchers conduct a set of analyses, including equivalence testing alongside bias testing (i.e., Bland-Altman), difference of means, mean absolute percentage error, and—whenever possible—epoch-by-epoch analyses.

4. Issues around the big picture of a phase-based validation process

4.1 There is a need for replication of validation studies.

We found that out of 391 wearables, 193 were included only once in a validation process. Additionally, only 25 devices were validated for both the children/adolescents and the adults target group across both laboratory and naturalistic conditions. In particular, we identified a lack of protocols conducted under naturalistic compared to laboratory conditions. The naturalistic part of the validation process encompasses short duration and spontaneous activities and behaviors performed in a way that is natural for the subject. By this, a more accurate representation of physical behavior during daily life is offered. Researchers should be aware that a wearable tested only under standardized laboratory conditions does not provide insights into the responsiveness of changes between physical behaviors.[37] Before using a wearable in applied studies, researchers may ensure that the wearable has been tested

under both laboratory and naturalistic conditions. Optimally, the results of those studies would be further replicated by an independent research team, resulting in 4 different high-quality validation studies.

4.2 Access to raw data is a key aspect of data reproducibility.

Only 31.1% of all studies reported the algorithm of the validated outcome (e.g., reported as a formula or at least provided further information). Moreover, most of those studies that reported algorithms validated research-grade devices and focused on biological state outcomes. A key quality aspect of reproducibility is providing transparent information about the devices used (i.e., all data processing steps, analytical approaches, and updates of algorithms). An example of variability in sensor specification is the diverse and non-harmonious ways of preprocessing data across the devices. In particular, different possibilities for signal processing raw data (e.g., rectifying, extracting low- and high-signal frequency) and transferring those signals into metrics are presented in the literature (e.g., counts,[38] movement acceleration intensity,[39] euclidian norm minus one,[40] or mean amplitude deviation[41]). A further notable challenge arises when algorithms are "black-boxed," rendering them inaccessible to the public, and researchers are unable to report more information, which is often the case when applying consumer-grade devices. This is not surprising since the primary aim of consumer device companies is to sell devices and keep the algorithm proprietary. However, this lack of information hinders the opportunity for refinement and optimization of the algorithms during the validation process. Optimally, researchers should have access to raw data or high-resolution feature data that are derived from the raw data. Furthermore, at the very least, comprehensive details of algorithms as well as the versions of the device and software should be made transparent. In this context and in line with the FAIR (Findability, Accessibility, Interoperability, and

Reusability) principle standards,[42] transparency is a requirement for reproducibility, comparability, stacking of datasets, or creating norms.

4.3 How to handle the release of a new version or series of a brand.

Across all studies, we found 391 devices from 166 different brands. Specifically, regarding fitness trackers and smartwatches such as Fitbit, we identified 18 series and models, highlighting the considerable challenges posed by the short life cycles of commercial-grade products. Notably, a new version of the Fitbit Inspire was released in less than 2 years. However, validating and adopting a new series of wearables in less than 2 years is deemed unrealistic, given the typically slower progress of research studies. As a potential solution, we refer to the prerequisite mentioned in point 2: Transparent provision of information by companies regarding all wearable and algorithm details. If the new series has not changed from the technical and algorithm side, we would suggest replicating the findings under naturalistic conditions and skipping the remaining phases of the validation framework. Challenges to and recommendations for improving validation processes and protocols of wearables to assess 24-h physical behavior are summarized in Fig. 4.

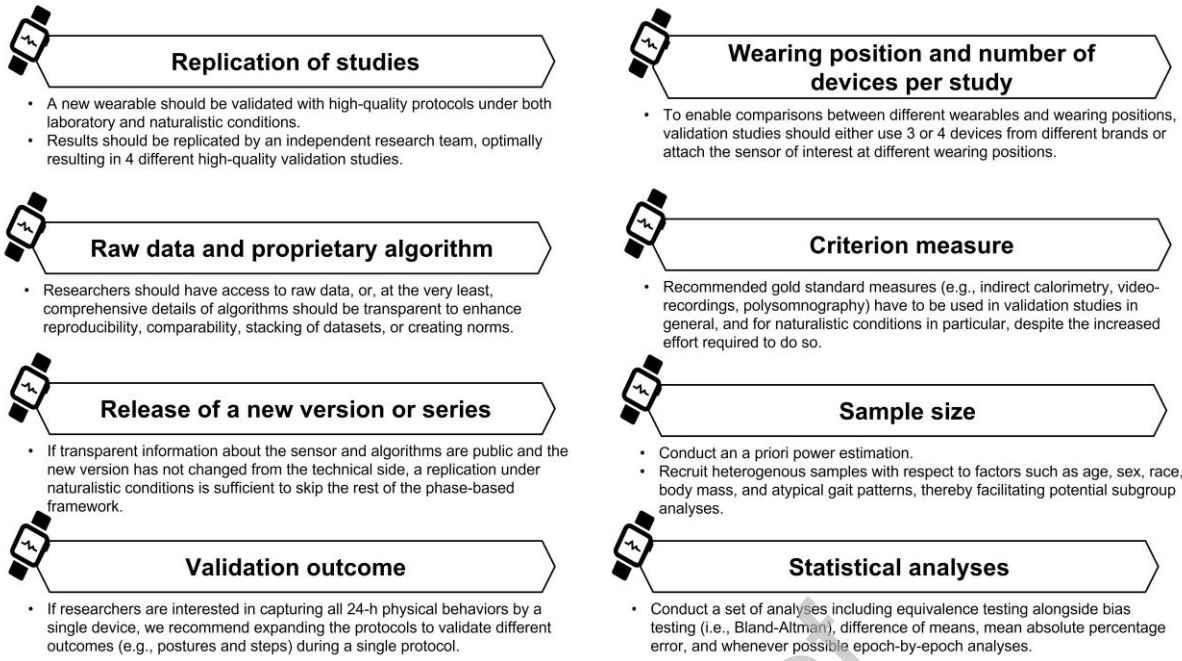

Fig. 4. Summary of key messages for the validation process and high-quality protocols.

5. A future scenario of wearables validation for the assessment of 24-h physical behavior

At the moment, no organization or institution is responsible for the regulatory mechanisms of the technical validation of wearables. Therefore, from our point of view, there is an imbalance between the important task of validation, the responsibility for it, and its allocated resources. For example, validation studies under naturalistic settings require the time-consuming process of labeling gold-standard criterion measures such as video recording. The launch of an open science initiative is a promising way to overcome this challenge.

Therefore, we launched the website http://www.wearable-landscape.info, which summarizes all validation protocols from our published reviews[1–4] and serves as a source of information on previous validation studies. The website will evolve into a living review, ensuring ongoing updates through periodic literature searches. We envision it becoming a central point for resource pooling within the scientific community. This could involve building a network for planning upcoming validation studies, where researchers can upload protocols and the community can receive alerts, fostering collaborations in resource-sharing, equipment

provision, studies conducted at various locations, and joint data analysis or pre-processing efforts.

Global trends of inactivity are projected to rise persistently over time,[43] heightening the relevance of 24-h physical behavior surveillance. Activity prescriptions by a physician or a healthcare system that integrates wearables (e.g., pay-as-you-live) is a conceivable reality. Fortunately, wearable technologies and their widespread accessibility to the public can help to tackle health issues. These devices have already transformed how we conduct research on, prescribe, and monitor 24-h physical behavior, offering the potential to advance and translate our understanding of how physical behaviors influence health conditions.[13] While we recognize great potential in wearable technology for the assessment of 24-h physical behavior, the rapid development and wide range of devices and algorithms should not overshoot rigorous technical validation, which is a critical prerequisite for expedient wearable application. In line with ongoing discussions on improving the validation and trustworthiness of digital health products, Mathews et al.[44] stress the need to distinguish the quality and value of digital health solutions. The call for a robust and transparent validation process for digital health products is directly applicable to the acquisition of 24-h physical behavior via wearables. Therefore, we encourage the scientific community and companies to collaborate, promoting openness and transparency with respect to the presentation of algorithms. This collaborative approach is essential to realize the full potential of open science toward valid wearables for everyone.

Authors' contributions

MG, BvHM, JF, SW, AB, SK, MR, UEP, and JB conceived, planned and drafted the manuscript. SKe, CK, CN, IT, MT, SS, KW, GM, CRN, and AW contributed to the intellectual development of this Viewpoint and provided critical feedback on the manuscript.

All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This research was funded in part by the German Research Foundation (Grant reference: 496846758].

References

1. 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 Exerc Med* 2022;8:e001267. doi:10.1136/bmjsem-2021-001267.

2. Giurgiu M, Nigg C, Fiedler J, et al. The assessment of 24-hour physical behavior in children and ddolescents via wearables: A systematic review of laboratory validation studies. *J Meas Phys Behav* 2022;5:263-76.

3. Giurgiu M, Ketelhut S, Kubica C, et al. Assessment of 24-hour physical behaviour in adults via wearables: A systematic review of validation studies under laboratory conditions. *Int J Behav Nutr Phys Act* 2023;20:68. doi:10.1186/s12966-023-01473-7.

4. Giurgiu M, Timm I, Becker M, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: Systematic review. *JMIR MHealth UHealth* 2022;10:e36377. doi:10.2196/36377.

5. Tremblay MS. Challenges in global surveillance of physical activity. *Lancet Child Adolesc Health* 2020;4:2-3.

6. Warburton DER, Katzmarzyk PT, Rhodes RE, Shephard RJ. Evidence-informed physical activity guidelines for Canadian adults. *Can J Public Health* 2007;98(Suppl. 2):S16-68.

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. doi:10.1136/bmjsem-2020-000874.

8. Ekelund U, Tarp J, Fagerland MW, et al. Joint associations of accelero-meter measured physical activity and sedentary time with all-cause mortality: A harmonised meta-analysis in more than 44,000 middle-aged and older individuals. *Br J Sports Med* 2020;54:1499-506.

9. Crowley P, Ikeda E, Islam SMS, et al. The surveillance of physical activity, sedentary behavior, and sleep: Protocol for the development and feasibility evaluation of a novel measurement system. *JMIR Res Protoc* 2022;11:e35697. doi:10.2196/35697.

10. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer perceptions of wearable Technology devices: Retrospective review and analysis. *JMIR MHealth UHealth* 2020;8:e17544. doi:10.2196/17544.

11. Fedor S, Lewis R, Pedrelli P, Mischoulon D, Curtiss J, Picard RW. Wearable technology in clinical practice for depressive disorder. *N Engl J Med* 2023;389:2457-66.

12. Kercher V, Kercher K, Levy P, et al. 2023 Fitness trends from around the Globe. *ACSMs Health Fit J* 2023;27:19-30.

13. Gill JM, Chico TJ, Doherty A, et al. Potential impact of wearables on physical activity guidelines and interventions: Opportunities and challenges. *Br J Sports Med* 2023;57:1223-5.

14. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65:2276-84.

15. Lee IM, Keadle SK, Matthews CE. Fitness trackers to guide advice on activity prescription. *JAMA* 2023;330:1733. doi:10.1001/jama.2023.19332.

16. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-recorded validation of wearable step counters under free-living conditions. *Med Sci Sports Exerc* 2018;50:1315-22.

17. Höchsmann C, Knaier R, Infanger D, Schmidt-Trucksäss A. Validity of smartphones and activity trackers to measure steps in a free-living setting over three consecutive days. *Physiol Meas* 2020;41:015001. doi:10.1088/1361-6579/ab635f.

18. Waddell A, Birkett S, Broom D, McGregor G, Harwood AE. Validating the Fitbit Charge 4© wearable activity monitor for use in physical activity interventions. *J Sci Med Sport* 2024;27:314-8.

19. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: A meta-analysis of 15 international cohorts. *Lancet Public Health* 2022;7:e219-28.

20. Brönneke JB, Müller J, Mouratis K, Hagen J, Stern AD. Regulatory, legal, and market aspects of smart wearables for cardiac monitoring. *Sensors* 2021;21:4937. doi:10.3390/s21144937.

21. EUR-Lex. *Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the regioNS eHealth Action Plan 2012–2020—Innovative healthcare for the 21st century.* Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0736. [accessed 18/4/2024]

22. DIME Society. *Digital Health Industry Regulatory Needs Assessment.* Available at: https://dimesociety.org/access-resources/dime-publications/#publications. [accessed 18/4/2024]

23. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. *Exerc Sport Sci Rev*. 2019;47:206-14.

24. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.

25. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc* 2012;44(Suppl. 1):S32-8.

26. 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.

27. Godfrey A, Conway R, Meagher D, ÓLaighin G. Direct measurement of human movement by accelerometry. *Med Eng Phys* 2008;30:1364-86.

28. Rosenberger ME, Fulton JE, Buman MP, et al. The 24-Hour Activity Cycle: A New Paradigm for Physical Activity. *Med Sci Sports Exerc* 2019;51:454-464. doi:10.1249/MSS.0000000000001811

29. Dumuid D, Pedišić Ž, Palarea-Albaladejo J, Martín-Fernández JA, Hron K, Olds T. Compositional data analysis in time-use epidemiology: What, why, how. *Int J Environ Res Public Health* 2020;17:2220. doi:10.3390/ijerph17072220

30. Argent R, Hetherington-Rauth M, Stang J, et al. Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure:

Expert statement and checklist of the INTERLIVE Network. *Sports Med* 2022;52:1817-32.

31. 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.

32. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. *Sleep Health* 2015;1:275-84.

33. Yuan H, Chan S, Creagh AP, et al. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *NPJ Digit Med* 2024;7:91. doi:10.1038/s41746-024-01062-3

34. O'Brien MW. Implications and recommendations for equivalence testing in measures of movement behaviors: A scoping review. *J Meas Phys Behav* 2021;4:353-62.

35. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. *Med Sci Sports Exerc* 2018;50:837-45.

36. O'Brien MW, Pellerine LP, Shivgulam ME, Kimmerly DS. Disagreements in physical activity monitor validation study guidelines create challenges in conducting validity studies. *Front Digit Health* 2023;4:1063324. doi:10.3389/fdgth.2022.1063324.

37. Clevenger KA, Montoye AHK. Systematic review of accelerometer responsiveness to change for measuring physical activity, sedentary behavior, or sleep. *J Meas Phys Behav* 2023;6:250-63.

38. Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors* 2010;10:7772-88.

39. van Someren EJ, Lazeron RH, Vonk BF, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: Implications for actigraphy in young and elderly subjects. *J Neurosci Methods* 1996;65:55-62.

40. Van Hees VT, Gorzelniak L, Dean León EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS ONE* 2013;8:e61691. doi:10.1371/journal.pone.0061691.

41. Vähä-Ypyä H, Vasankari T, Husu P, Suni J, Sievänen H. A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. *Clin Physiol Funct Imaging* 2015;35:64-70.

42. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18.

43. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1.6 million participants. *Lancet Child Adolesc Health* 2020;4:23-35.

44. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: A path to validation. *NPJ Digit Med* 2019;2:38. doi:10.1038/s41746-019-0111-3.