

Viewpoint

# Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity

Peter Düking<sup>1,2\*</sup>, MSc; Franz Konstantin Fuss<sup>3\*</sup>, MD, PhD; Hans-Christer Holmberg<sup>2,4,5\*</sup>, PhD; Billy Sperlich<sup>1\*</sup>, PhD

<sup>1</sup>Integrative & Experimental Exercise Science & Training, Institute for Sport Sciences, University of Würzburg, Würzburg, Germany

<sup>2</sup>Swedish Winter Sports Research Centre, Mid Sweden University, Östersund, Sweden

<sup>3</sup>Smart Equipment Engineering and Wearable Technology Research Program, Centre for Design Innovation, Swinburne University of Technology, Melbourne, Australia

<sup>4</sup>School of Sport Sciences, University of Tromsø - The Arctic University of Norway, Tromsø, Norway

<sup>5</sup>School of Kinesiology, University of British Columbia, Vancouver, BC, Canada

\* all authors contributed equally

**Corresponding Author:**

Peter Düking, MSc

Integrative & Experimental Exercise Science & Training

Institute for Sport Sciences

University of Würzburg

Judenbühlweg 11

Würzburg, 97082

Germany

Phone: 49 931 31 ext 8479

Email: [peterdueking@gmx.de](mailto:peterdueking@gmx.de)

## Abstract

Although it is becoming increasingly popular to monitor parameters related to training, recovery, and health with wearable sensor technology (wearables), scientific evaluation of the reliability, sensitivity, and validity of such data is limited and, where available, has involved a wide variety of approaches. To improve the trustworthiness of data collected by wearables and facilitate comparisons, we have outlined recommendations for standardized evaluation. We discuss the wearable devices themselves, as well as experimental and statistical considerations. Adherence to these recommendations should be beneficial not only for the individual, but also for regulatory organizations and insurance companies.

(*JMIR Mhealth Uhealth* 2018;6(4):e102) doi:[10.2196/mhealth.9341](https://doi.org/10.2196/mhealth.9341)

**KEYWORDS**

activity tracker; data mining; Internet of Things; load management; physical activity; smartwatch

## Introduction

Wearable sensors (so-called “wearables”) are currently the world’s leading trend in fitness [1,2] and are being employed widely by various groups to monitor variables related to health, physical activity, training load, and recovery [3,4], often with the goal of individualizing physical activity and improving performance. Several insurance companies promote such monitoring [5] and an increasing number of organizations that regulate sports (eg, the International Football Association Board [6]), allow wearables to be worn during competitions (albeit with certain limitations).

If wearables are to be of value in enhancing health and performance [4], it is becoming more and more imperative that the data they supply are proven to be trustworthy by employing scientific approaches [7]. Unfortunately, wearables are often marketed with aggressive and exaggerated claims that lack a sound scientific basis [7], and the unreliable data they provide (and/or interpretation thereof) has resulted in costly class-action lawsuits [8] and provides little or no value to the customer.

Recent scientific evaluation of wearable data has involved widely heterogeneous study designs (including the nature and size of the study population), methodologies, criteria for comparisons, terminologies, and statistical analyses, as well as varying intensities/modalities of exercise. Assessment of novel

technology may be influenced by the particular test conditions employed [9]. For example, laboratory data may not be transferable to real-life situations and data trustworthy in a resting condition or during low-intensity exercise may become less valid at higher intensity (eg, due to motion artifacts). Thus, variations in methodology complicate the comparison of scientific evaluations of wearable data.

From our perspective, athletes, manufacturers of wearables, and organizations concerned with health, sports, and insurance could all benefit from basic recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors. The aim of this paper is to formulate such recommendations.

## *Factors Inherent to the Wearables Themselves*

### **Sensor Characteristics**

Wearables contain a wide variety of sensors (eg, electrochemical, optical, acoustic, and/or pressure-sensitive), as well as inertial measurement units and global navigation satellite systems (including global positioning systems [GPS]). More than one of these are often present within the same device. These sensors, produced by various manufacturers, are designed to monitor a variety of internal (eg, heart rate, tissue oxygenation, distribution of plantar pressure) and/or external (eg, acceleration of body segments, speed while exercising) parameters, mostly noninvasively [3]. With multi-sensor devices, the quality of data and parameters derived depends on the interplay between the sensors, each of which must therefore be scrutinized both independently and in combination with the others. Consideration of individual sensors is beyond the scope of the present recommendations and we refer the reader to other relevant work for such information [3,10,11].

### **Software**

The nature of the software in the wearable itself, as well as of the software in any accompanying device (ie, laptop, smartphone application) exerts a considerable influence on data quality. For example, the software in GPS receivers or analytical software on an accompanying device may actually alter data [12-14]. We therefore urge researchers to describe the software utilized by the wearable and accompanying devices and/or the involvement of “cloud” technology in detail.

### **Acquisition of Raw Data: Sampling Frequency and Filtering**

Although of less concern to the private consumer, to improve the reliability, sensitivity, and validity of data used for research purposes we recommend that manufacturers provide access to raw data. This issue is of particular interest in the case of multi-sensor devices, which often calculate a single value by combining data from several sensors (a common example being calculation of energy expenditure by merging heart rate with several GPS parameters), yet the contribution by each individual sensor is often unclear. Describing these contributions could enhance scientific trustworthiness (eg, by improving the algorithms employed).

A high sampling frequency, which normally enhances data quality, may be achieved artificially by filtering techniques (eg, interpolation) that produce no actual improvement in this quality [15]. Consequently, both the sampling frequency and any filtering techniques applied should be described in detail.

### **Durability**

Sensors can deteriorate or even wear out with extended use and it is clearly important to describe the durability of the wearable and its sensor technology, at least as indicated by the manufacturer. Unlike laboratory equipment, most wearables are not checked routinely, making such description essential. Wearable devices are typically brand-new when evaluated and the quality and trustworthiness of the data they provide may change with use.

### **Precise Reporting of Anatomical Positioning**

Wearables and their algorithms are often designed for use at a specific position or region of the body, which, consequently, must be indicated clearly. In certain cases, imprecise positioning may attenuate data quality [3]. For example, sensors for surface electromyography incorporated into clothing must be positioned precisely on the muscle, preferably along the midline, halfway between the entrance of the nerve and myotendinous junction [16]. On a daily basis, such accurate positioning may prove difficult, especially since this is often performed by nonprofessionals. Moreover, signal reproducibility may be affected by repeated donning and removal of garments. Consequently, we encourage researchers to describe in detail the positioning of wearables, as well as reproducibility of data. Researchers often evaluate several wearables simultaneously and such devices in close proximity can interfere with one another [15]. We recommend strongly that any potential interference be controlled for.

## *Experimental Considerations*

### **Study Population**

Selection of the study population (eg, cyclists, runners or team members, elite or recreational athletes, youth or adults, men or women) should accurately reflect the intended use of the wearable. Each population behaves differently (eg, with respect to lifestyle) and algorithms should be transferred from one specific population to another only with great care. The inclusion and exclusion criteria for participants must be described clearly. If anyone opts out of the experimental procedure or data analysis, a reason should be given.

### **Exercise Protocol**

The intended purpose and conditions for use of the wearable should be clarified. If designed for monitoring general activity, data should be collected in connection with various forms of exercise (eg, running, cycling, rowing, intermittent activities, activities of daily living) of varying intensity (eg, resting, submaximal, high), in different positions (lying, sitting, or standing), and/or while moving freely. If a wearable is intended to be used in connection with team sports such as soccer, a protocol mimicking the demands of this sport—including low-speed running, straight sprints, change-of-direction, and

tackling—is much more preferable than running constantly at low speed only.

### Potential Confounders

Factors that could influence the outcome, such as temperature and humidity, the warm-up procedure, nutritional status, and any form of encouragement, should resemble the real-life situation as closely as possible [17,18] and be described in detail.

Other potential confounders may also need to be taken into consideration. For example, sensors that monitor electrical signals (eg, for electromyography or electrocardiography) may be influenced by other devices, such as a participant's pacemaker. Optical sensors (eg, for photoplethysmography) can be affected by the photosensitivity of the skin or by vasoconstriction [19,20]. In the case of GPS receivers, the horizontal dilution of precision, as well as the number of satellites to which the wearable is connected, should be reported [15]. Although there are no clear rules, two wearables should not be tested at the same time (eg, one on top of the other) or, if they are, potential interference and crosstalk should be examined for by switching the positions of the devices [21]. Adequate controlling for numerous confounding factors requires a good understanding of both the sensor technology and associated physiological and/or biomechanical processes.

### Special Considerations Concerning Reliability

Intradvice reliability concerns reproducibility within the same device [22,23], while interdevice reliability (reproducibility with different devices) is to be tested if the devices in question are intended for interchangeable use [12]. Both types of reliability should be confirmed routinely. Recently, it has been recommended that at least 50 participants and three trials should be involved in order to obtain precise estimates of reliability [23]. When multiple trials are performed at different times, potential confounders must vary as little as possible.

### Special Considerations Concerning Validity

Several different types of validity (eg, logical, convergent, and construct validity [24,25]) are probably equally important in this context, but discussion of these in detail is beyond the present scope and we refer the interested reader to other relevant articles [24,25]. Here, we focus on concurrent criterion validity, since this is probably easiest to access with respect to wearables. Concurrent criterion validity evaluates the association between data provided by the new device and another device considered to be more valid (sometimes referred to as a criterion measure or “gold-standard”) [23,25].

For certain parameters, there are generally-accepted criterion measures (eg, polysomnographic parameters of sleep [26] and an ingestible telemetric sensor for core body temperature [27]). However, for others (eg, energy expenditure at several timepoints while moving freely and in-shoe plantar pressure) no such measures are currently available. We encourage researchers to describe the trustworthiness of their criterion measures and strongly discourage the use of measures not considered to be “gold-standard” for validation of the quality of wearable data.

## Statistical Analyses

### Overview

The various statistical approaches for evaluating the reliability or validity of wearables all have limitations [28,29]. Without discouraging the usage of other robust approaches (eg, the Standard Error of Measurement for reliability studies [28] or Bland-Altman plots for validity studies [30,31]), we propose one possible approach to statistical assessment of wearable data concerning reliability, sensitivity, and validity in the following sections.

### Reliability

Reliability should be documented in terms of intrasubject variability (eg, measured as standard deviation, “...the random variation in a measure when one individual is tested many times”), which is possibly the most important indicator of the reliability of measures of performance and sometimes referred to as typical error (TE) [23]. The TE can also be expressed as the coefficient of variation (%CV) [23] and we encourage the reporting of both.

Another measure of reliability (eg, “...the change in mean value between 2 trials...”) assesses systematic bias in combination with random variations [23]. The random variation is simply a sampling error, which tends to be smaller with larger samples. Systematic bias can be due to learning by (and training of) subjects or effects related to fatigue, and consequently can often be minimized by familiarization trials or adequate rest between trials, respectively [23].

In addition, researchers should assess test-retest reliability with the intraclass correlation coefficient [32], which “represents how closely the values of one trial track the values of another as we move our attention from individual to individual” [23] or, in other words, the reliability “of the position or rank of individuals in the group relative to others” [28]. Moreover, to determine whether data provided by different wearables can be used interchangeably, it may be of interest to evaluate interdevice reliability, previously accomplished by calculating the %CV between the devices when worn simultaneously [12].

### Sensitivity

Wearables designed to track changes in performance and/or parameters over time must, of course, be sensitive to such changes [33]. Even with a reliable test, the noise can be high enough to mask changes in parameters [33]. In the case of individual elite athletes, for whom certain fitness parameters are directly correlated with performance (eg, energy expenditure at a given running intensity; the lower, the less intense), the smallest worthwhile change (SWC) is 30% of the individual's typical variation in performance [34]. Where there is no clear relationship between parameters of fitness and performance (eg, strength and team sport performance), it has been proposed that the SWC be calculated (0.2 times the between-subject standard deviation, based on Cohen's effect size principle) and compared with the noise of the measuring device or test [33,34]. This noise can be expressed as the TE, which can be obtained from reliability studies, as described above. A TE less than, similar to, or higher than the SWC can be rated as “good,” “OK,” or

“marginal,” respectively [33]. When assessing sensitivity, similar and reliable experimental approaches are required.

### Validity

Linear regression analysis can be employed to identify bias and provide an estimate of the TE in wearable data [29,35,36]. Furthermore, Pearson’s product-moment correlation coefficient should be calculated [36] to compare the degree of association [33,37] between data obtained with the criterion measure and the wearable. However, a significant correlation does not definitively mean that these data do not differ and is not, therefore, on its own a sufficient indicator of validity [30].

### Conclusions

Here, we have outlined general recommendations (summarized in Table 1) for the evaluation of the trustworthiness of

monitoring training load, recovery, and health by wearables. We are well aware that with certain technologies, other methodological considerations may be of particular importance and that new approaches are emerging constantly. Although evaluation may not be possible or even desirable in every individual context, findings in one situation should be transferred to another only with great care and appropriate justification.

The market for wearables is growing exponentially and their scientific evaluation in a trustworthy manner needs to keep pace. The success of a wearable device depends on gaining the trust of the consumer, stakeholders, and policymakers alike (eg, by transparent reporting of standardized validation, ideally carried out by an independent research institution). We are convinced that these recommendations can aid manufacturers of wearables, athletes, coaches, team managers, insurance companies, and other stakeholders and policymakers alike in evaluating wearable sensor technologies and/or selecting appropriate devices.

**Table 1.** Checklist of important considerations associated with the evaluation of data provided by wearables.

Factor	Action/recommendation
Sensor characteristics	<ul style="list-style-type: none"> <li>• Scrutiny of each sensor</li> </ul>
Software	<ul style="list-style-type: none"> <li>• Specify calculations/algorithms</li> <li>• Report the version of software and firmware involved</li> </ul>
Raw data	<ul style="list-style-type: none"> <li>• Report sampling frequency</li> <li>• Report filtering techniques and aggregation</li> </ul>
Durability	<ul style="list-style-type: none"> <li>• Report the durability and age of the device</li> </ul>
Anatomical positioning	<ul style="list-style-type: none"> <li>• Report the precise anatomical positioning of sensors</li> <li>• Report signal reproducibility upon repeated putting on and taking off</li> <li>• Report considerations concerning positioning</li> <li>• Control for and describe potential interference</li> </ul>
Study population	<ul style="list-style-type: none"> <li>• Describe the target population</li> <li>• Specify inclusion and exclusion criteria</li> <li>• Generalize to other populations only with great care</li> </ul>
Exercise protocol	<ul style="list-style-type: none"> <li>• Describe conditions (eg, ambient temperature, altitude) in as much detail as possible</li> <li>• Investigate different forms of exercise (running, cycling, walking, moving freely)</li> <li>• Apply different intensities (lying, sitting, low and high intensity)</li> </ul>
Confounders	<ul style="list-style-type: none"> <li>• Report any potential confounding factors</li> <li>• Perform assessment in both controlled and real-life scenarios</li> <li>• Check for potential crosstalk between devices</li> </ul>
Assessment of reliability	<ul style="list-style-type: none"> <li>• Determine intradevice and interdevice reliability</li> <li>• Document intrasubject standard deviation</li> <li>• Report the coefficient of variation</li> <li>• Calculate the intraclass correlation coefficient</li> <li>• Recruit at least 50 participants</li> <li>• Report systematic bias</li> </ul>
Assessment of sensitivity	<ul style="list-style-type: none"> <li>• Calculate the smallest worthwhile change</li> </ul>
Assessment of validity	<ul style="list-style-type: none"> <li>• Choose an appropriate criterion measure and assess the reliability of this measure as well</li> <li>• Perform linear regression analysis</li> <li>• Calculate Pearson’s product-moment correlation</li> </ul>

## Acknowledgments

This publication was funded by the German Research Foundation (DFG) and the University of Wuerzburg through the Open Access Publishing funding program.

## Conflicts of Interest

None declared.

## References

1. Thompson WR. Worldwide survey of fitness trends for 2017. *ACSM Health Fitness J* 2016;20(6):8-17. [doi: [10.1249/FIT.0000000000000252](https://doi.org/10.1249/FIT.0000000000000252)]
2. Thompson WR. Worldwide survey of fitness trends for 2016: 10th anniversary edition. *ACSM Health Fitness J* 2015;19(6):9-18. [doi: [10.1249/FIT.0000000000000164](https://doi.org/10.1249/FIT.0000000000000164)]
3. Düking P, Hotho A, Holmberg H, Fuss FK, Sperlich B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Front Physiol* 2016;7:71 [FREE Full text] [doi: [10.3389/fphys.2016.00071](https://doi.org/10.3389/fphys.2016.00071)] [Medline: [27014077](https://pubmed.ncbi.nlm.nih.gov/27014077/)]
4. Düking P, Holmberg H, Sperlich B. Instant biofeedback provided by wearable sensor technology can help to optimize exercise and prevent injury and overuse. *Front Physiol* 2017 Apr;8:167 [FREE Full text] [doi: [10.3389/fphys.2017.00167](https://doi.org/10.3389/fphys.2017.00167)] [Medline: [28420998](https://pubmed.ncbi.nlm.nih.gov/28420998/)]
5. Baas J. Die Techniker. 2016 Dec 22. Digitalisierung nutzen, um Gesundheit und die Solidargemeinschaft zu fördern URL: <https://www.tk.de/tk/themen/digitale-gesundheit/gesundheitsfoederung-durch-fitnesstracker-interview-dr-jens-baas/931248> [accessed 2018-04-04] [WebCite Cache ID [6yQE4yKX4](https://www.webcitation.org/6yQE4yKX4)]
6. Brud L. The International Football Association Board. 2015 May. Amendments to the laws of the game - 2015/2016 and information on the completed reform of The International Football Association Board URL: [http://resources.fifa.com/mm/document/affederation/ifab/02/60/91/38/circular\\_log\\_amendments\\_2015\\_v1.0\\_en\\_neutral.pdf](http://resources.fifa.com/mm/document/affederation/ifab/02/60/91/38/circular_log_amendments_2015_v1.0_en_neutral.pdf) [accessed 2017-02-19] [WebCite Cache ID [6yQEWMtL](https://www.webcitation.org/6yQEWMtL)]
7. Sperlich B, Holmberg H. Wearable, yes, but able...?: it is time for evidence-based marketing claims!. *Br J Sports Med* 2016 Dec 16 [FREE Full text] [doi: [10.1136/bjsports-2016-097295](https://doi.org/10.1136/bjsports-2016-097295)] [Medline: [27986762](https://pubmed.ncbi.nlm.nih.gov/27986762/)]
8. Eadicicco L. Time. 2016 May 23. 4 things to know about the Fitbit accuracy lawsuit URL: <http://time.com/4344675/fitbit-lawsuit-heart-rate-accuracy/> [accessed 2018-04-04] [WebCite Cache ID [6yQIfOxnF](https://www.webcitation.org/6yQIfOxnF)]
9. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc* 2012 Jan;44(1 Suppl 1):S32-S38 [FREE Full text] [doi: [10.1249/MSS.0b013e3182399cf7](https://doi.org/10.1249/MSS.0b013e3182399cf7)] [Medline: [22157772](https://pubmed.ncbi.nlm.nih.gov/22157772/)]
10. Bandodkar AJ, Wang J. Non-invasive wearable electrochemical sensors: a review. *Trends Biotechnol* 2014 Jul;32(7):363-371. [doi: [10.1016/j.tibtech.2014.04.005](https://doi.org/10.1016/j.tibtech.2014.04.005)] [Medline: [24853270](https://pubmed.ncbi.nlm.nih.gov/24853270/)]
11. Tiwana MI, Redmond SJ, Lovell NH. A review of tactile sensing technologies with applications in biomedical engineering. *Sensor Actuat A-Phys* 2012 Jun;179(1):17-31. [doi: [10.1016/j.sna.2012.02.051](https://doi.org/10.1016/j.sna.2012.02.051)]
12. Buchheit M, Al HH, Simpson BM, Palazzi D, Bourdon PC, Di Salvo V, et al. Monitoring accelerations with GPS in football: time to slow down? *Int J Sports Physiol Perform* 2014 May;9(3):442-445. [doi: [10.1123/ijsspp.2013-0187](https://doi.org/10.1123/ijsspp.2013-0187)] [Medline: [23916989](https://pubmed.ncbi.nlm.nih.gov/23916989/)]
13. Roe G, Darrall-Jones J, Black C, Shaw W, Till K, Jones B. Validity of 10-HZ GPS and timing gates for assessing maximum velocity in professional rugby union players. *Int J Sports Physiol Perform* 2017 Jul;12(6):836-839. [doi: [10.1123/ijsspp.2016-0256](https://doi.org/10.1123/ijsspp.2016-0256)] [Medline: [27736256](https://pubmed.ncbi.nlm.nih.gov/27736256/)]
14. Lee J, Kim Y, Bai Y, Gaesser GA, Welk GJ. Validation of the SenseWear mini armband in children during semi-structure activity settings. *J Sci Med Sport* 2016 Jan;19(1):41-45. [doi: [10.1016/j.jsams.2014.10.004](https://doi.org/10.1016/j.jsams.2014.10.004)] [Medline: [25459233](https://pubmed.ncbi.nlm.nih.gov/25459233/)]
15. Malone JJ, Lovell R, Varley MC, Coutts AJ. Unpacking the black box: applications and considerations for using GPS devices in sport. *Int J Sports Physiol Perform* 2017 Apr;12(Suppl 2):S218-S226. [doi: [10.1123/ijsspp.2016-0236](https://doi.org/10.1123/ijsspp.2016-0236)] [Medline: [27736244](https://pubmed.ncbi.nlm.nih.gov/27736244/)]
16. De Luca CJ. The use of surface electromyography in biomechanics. *J Appl Biomech* 1997;13(2):135-163.
17. Halperin I, Pyne DB, Martin DT. Threats to internal validity in exercise science: a review of overlooked confounding variables. *Int J Sports Physiol Perform* 2015 Oct;10(7):823-829. [doi: [10.1123/ijsspp.2014-0566](https://doi.org/10.1123/ijsspp.2014-0566)] [Medline: [25756869](https://pubmed.ncbi.nlm.nih.gov/25756869/)]
18. Hopkins WG, Schabert EJ, Hawley JA. Reliability of power in physical performance tests. *Sports Med* 2001;31(3):211-234. [Medline: [11286357](https://pubmed.ncbi.nlm.nih.gov/11286357/)]
19. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *J Med Eng Technol* 2015;39(5):264-271. [doi: [10.3109/03091902.2015.1047536](https://doi.org/10.3109/03091902.2015.1047536)] [Medline: [26112379](https://pubmed.ncbi.nlm.nih.gov/26112379/)]
20. Chan ED, Chan MM, Chan MM. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respir Med* 2013 Jun;107(6):789-799 [FREE Full text] [doi: [10.1016/j.rmed.2013.02.004](https://doi.org/10.1016/j.rmed.2013.02.004)] [Medline: [23490227](https://pubmed.ncbi.nlm.nih.gov/23490227/)]
21. Weizman Y, Tan A, Fuss FK. Benchmarking study of smart-insole forces and centre of pressure. Submitted .
22. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998 Oct;26(4):217-238. [Medline: [9820922](https://pubmed.ncbi.nlm.nih.gov/9820922/)]

23. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000 Jul;30(1):1-15. [Medline: [10907753](#)]
24. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: convergent validity. *Sports Med* 2002;32(12):795-808. [Medline: [12238942](#)]
25. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports Med* 2008;38(4):297-316. [Medline: [18348590](#)]
26. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003 May 01;26(3):342-392. [Medline: [12749557](#)]
27. Byrne C, Lim CL. The ingestible telemetric body core temperature sensor: a review of validity and exercise applications. *Br J Sports Med* 2007 Mar 01;41(3):126-133. [doi: [10.1136/bjsm.2006.026344](#)]
28. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005 Feb;19(1):231-240. [doi: [10.1519/15184.1](#)] [Medline: [15705040](#)]
29. Hopkins WG. Sport Science. 2004. Bias in Bland-Altman but not regression validity analyses URL: <http://www.sportsci.org/jour/04/wghbias.htm> [accessed 2018-04-04] [WebCite Cache ID 6yQIpBO2C]
30. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Med Sci Sports Exerc* 2012 Jan;44(1 Suppl 1):S39-S49. [doi: [10.1249/MSS.0b013e3182399d8f](#)] [Medline: [22157773](#)]
31. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999 Jun;8(2):135-160. [Medline: [10501650](#)]
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979 Mar;86(2):420-428. [Medline: [18839484](#)]
33. Buchheit M, Lefebvre B, Laursen PB, Ahmaidi S. Reliability, usefulness, and validity of the 30-15 Intermittent Ice Test in young elite ice hockey players. *J Strength Cond Res* 2011 May;25(5):1457-1464. [doi: [10.1519/JSC.0b013e3181d686b7](#)] [Medline: [21522077](#)]
34. Hopkins WG. Sport Science. 2004. How to interpret changes in an athletic performance test URL: <http://www.sportsci.org/jour/04/wghtests.htm> [accessed 2018-04-04] [WebCite Cache ID 6yQIFS8YY]
35. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc* 2009 Jan;41(1):3-13. [doi: [10.1249/MSS.0b013e31818cb278](#)] [Medline: [19092709](#)]
36. Hopkins WG. Sport Science. 2010. A socratic dialogue on comparison of measures URL: <http://www.sportsci.org/2010/wghmeasures.htm> [accessed 2018-04-04] [WebCite Cache ID 6yQIP5eBC]
37. McCall A, Fanchini M, Coutts AJ. The modern-day sport-science and sports-medicine “quest for the holy grail”. *Int J Sports Physiol Perform* 2017 May;12(5):704-706. [doi: [10.1123/ijsp.2017-0137](#)] [Medline: [28488907](#)]

## Abbreviations

- %CV:** coefficient of variation
- GPS:** global positioning system
- SWC:** smallest worthwhile change
- TE:** typical error

*Edited by G Eysenbach; submitted 06.11.17; peer-reviewed by L Ardigo, S Trost; comments to author 18.12.17; revised version received 08.02.18; accepted 17.02.18; published 30.04.18*

*Please cite as:*

*Düking P, Fuss FK, Holmberg HC, Sperlich B*

*Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity*

*JMIR Mhealth Uhealth* 2018;6(4):e102

URL: <http://mhealth.jmir.org/2018/4/e102/>

doi: [10.2196/mhealth.9341](#)

PMID: [29712629](#)

©Peter Düking, Franz Konstantin Fuss, Hans-Christer Holmberg, Billy Sperlich. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 30.04.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.