

Viewpoint

Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations

Francesco Sartor¹, PhD; Gabriele Papini², MSc; Lieke Gertruda Elisabeth Cox¹, PhD; John Cleland³, MD, FRCP

¹Personal Health, Philips Research, Eindhoven, Netherlands

²Department of Electrical Engineering, Technical University Eindhoven, Eindhoven, Netherlands

³Institute of Health & Well-Being, Robertson Centre for Biostatistics and Clinical Trials, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:

Francesco Sartor, PhD

Personal Health

Philips Research

High Tech Campus

Eindhoven,

Netherlands

Phone: 31 61 550 9627

Email: francesco.sartor@philips.com

Abstract

Wearable sensor technology could have an important role for clinical research and in delivering health care. Accordingly, such technology should undergo rigorous evaluation prior to market launch, and its performance should be supported by evidence-based marketing claims. Many studies have been published attempting to validate wrist-worn photoplethysmography (PPG)-based heart rate monitoring devices, but their contrasting results question the utility of this technology. The reason why many validations did not provide conclusive evidence of the validity of wrist-worn PPG-based heart rate monitoring devices is mostly methodological. The validation strategy should consider the nature of data provided by both the investigational and reference devices. There should be uniformity in the statistical approach to the analyses employed in these validation studies. The investigators should test the technology in the population of interest and in a setting appropriate for intended use. Device industries and the scientific community require robust standards for the validation of new wearable sensor technology.

(*J Med Internet Res* 2018;20(7):e10108) doi: [10.2196/10108](https://doi.org/10.2196/10108)

KEYWORDS

sensor technology; accuracy; wearable; telemonitoring

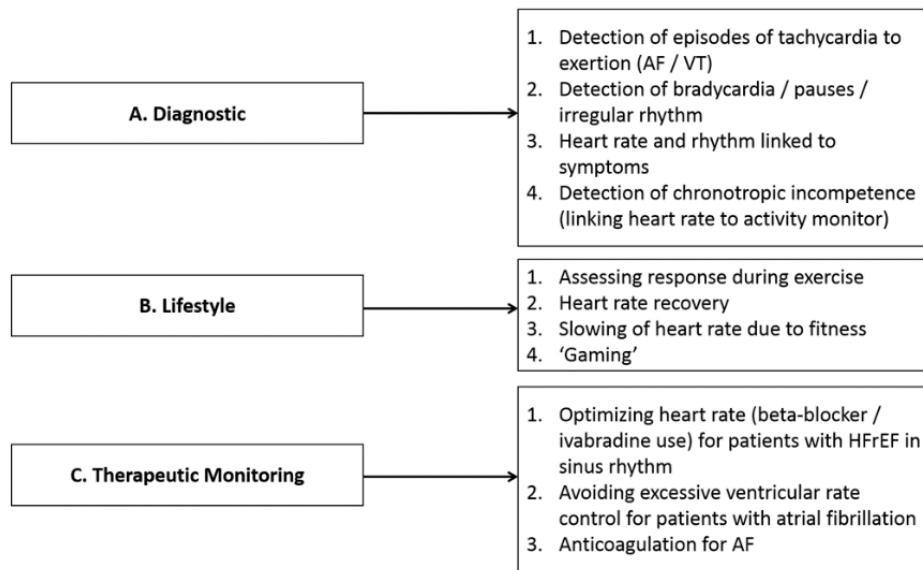
In the past 5 years, there has been a huge proliferation of wrist-worn heart rate monitors, often embedded in smart-bands and smartwatches, which can generate a vast amount of data on lifestyle, physiology, and disease providing exciting opportunities for future health applications. Wearable sensor technology could have an important role for clinical research and in delivering health care [1]. Wearable sensors can be used to encourage healthier living (possible delaying or preventing the onset of disease), screen for incident disease, and provide unobtrusive continuous monitoring for people with chronic illnesses in order to optimize care and detect disease progression and complications. In [Figure 1](#), we show an overview of potential continuous heart rate monitoring applications. New diagnostic applications could become possible thanks to the integration of heart rate and personal information such as age, sex, fitness, activity type, and symptoms. A large number of lifestyle apps and games are emerging thanks to continuous

heart rate monitoring, currently most of them related to fitness (eg, Google Fit, Strava) or biofeedback relaxation (eg, Letter Zap, Skip a Beat). It is conceivable that health-promoting apps or games based on heart rate will soon be developed. Wearable heart rate monitors could also enable therapeutic monitoring such as medication titration. Accordingly, such monitors should undergo rigorous evaluation prior to market launch, and their performance should be supported by evidence-based marketing claims [1].

There are several types of validation studies. These studies may be marketing claim validations or medical claim validations for medical grade certification. They are usually done by the manufacturers, sometimes in collaboration with clinical sites, on unreleased products. There may also be benchmarking validation studies, where several commercially available competing products are compared to one another and against a

reference. In some cases, there may be even single device validation studies.

Figure 1. Brief overview of potential clinical and nonclinical applications derivable from continuous heart rate monitoring. AF/VT: atrial fibrillation/ventricular tachycardia; HFrEF: heart failure with reduced ejection fraction.



The latter 2 types are generally performed by academic or clinical centers even though industries often engage in such comparisons as well. The only studies which go through a strict quality regulatory framework are medical claim validation studies for medical grade certification (eg, Food and Drug Administration in the United States, medical CE [Conformité Européene] marking in Europe) [2,3]. As a consequence, many nonmedical devices are released on the market without rigorous validation.

In Europe, the choice on how to position a device is the responsibility of the manufacturer, whereas in the United States, this decision can be overruled if the device is perceived to have potential health risks for the user [4]. Because manufacturers can decide whether or not they wish to comply with medical certification regulations, this inevitably leads to heterogeneity in what validations are done. In our view, the lack of stringent regulations for the release of nonmedical heart rate monitoring devices should not justify the lack of standard requirements for validating this technology. The adoption of such technology by health care professionals could be hampered by their liability in case of adverse events when using commercially available nonmedical devices. The authors of this viewpoint agree with Quinn [4], who suggests “a more pragmatic, risk-based approach,” which takes a case-by-case look at commercial solutions that may or may not meet the standards required of medical devices. This approach should be applied to promote technology adoption and at the same time safeguard the safety of end-users. Here, we give an overview of clinical applications exploiting wearable heart rate monitors.

In a Research Letter recently published in JAMA [5], the performance of several commercially available, wrist-worn photoplethysmography (PPG)-based heart rate monitors was reported. The authors concluded that PPG-based monitoring was not suitable “when accurate measurement of heart rate is imperative.” The authors of that Research Letter acknowledged their report had limitations, including testing only 1 type of

activity (treadmill), only in healthy people, and noncontinuous monitoring. Many other studies have been published validating wrist-worn PPG-based heart rate monitoring devices [6-14] but fail to show consensus in favor of or against the accuracy of this sensing technology.

The authors believe that the reason why many validations did not provide conclusive evidence of the validity of wrist-worn PPG-based heart rate monitoring devices is mostly methodological. Studies conducted by teams with a biomedical engineering background are more concerned with addressing problems like signal synchronization and averaging, while research teams with a sports medicine background are more concerned with target groups and exercise protocols. Moreover, clinicians are primarily interested in apps related to telemonitoring, in-hospital or remote. Each approach has its methodological shortcomings. The aim of this viewpoint is to suggest a more consistent and robust approach to validating monitoring technologies.

When validating heart rate monitoring devices, it is sensible to follow a common definition of accuracy. The American National Standards Institute standard for cardiac monitors, heart rate meters, and alarms defines accuracy as a “readout error of no greater than ±10% of the input rate or ±5 bpm, whichever is greater” [15]. Once accurate heart rate is defined, it is also good to agree on what to use as a gold standard. Electrocardiography (ECG) is the accepted gold standard for heart rate monitoring. Nevertheless, ECG, as with PPG, can be severely affected by artifacts [16]. Yet it is generally accepted that PPG-based heart rate monitoring suffers from inherent drawbacks (eg, more difficult peak detection, higher sensitivity to motion artifacts) compared to ECG-based monitoring [16].

The validation strategy should consider the nature of data provided by investigational devices (ID) and reference devices (RD). Heart rate values are always derived from more complex signals (eg, ECG, PPG). Thus, even when the ID and RD have

the same output rate (eg, 1 heart rate value per second) and these outputs are well synchronized, the beats compared may not belong to the same time intervals. The method used to extract information from the raw data (eg, time domain or frequency domain) and the averaging strategy (eg, interbeat intervals or 5-second periods) of the raw data will determine a specific time lag for each heart rate value. Ideally, researchers should have access to the raw data. This is often not possible, and it should be acknowledged as a limitation.

Researchers should realize that their RD (often an ECG device) will not always be accurate. Unless there is a quality check on the validity of the ECG, a second reference device should be used such as a second ECG-based sensor applied in a different

manner (eg, patch versus chest strap) and using a different software algorithm for calculating heart rate. When the two RDs fail to agree, no comparison should be made between RD and ID outputs (Figure 2). As mentioned earlier, even the RD (for example ECG patch or ECG strap) in certain circumstances may suffer from inaccuracy due to artifacts (eg, motion artifacts). Based on our own experience in testing hundreds of subjects, we realized that ECG patches perform particularly badly when the skin under the electrodes is stretched or excessively wet. ECG straps perform rather poorly when the skin gets too dry, the strap loosens up, and for certain anatomical shapes (pectus excavatum). These problems must be reported by the researcher.

Figure 2. Correlation between 3 heart rate (HR) monitoring devices and the electrocardiography (ECG) reference. When the 2 chest straps and the wrist-worn photoplethysmography (PPG) heart rate monitors consistently disagree with the reference, their points depart from the 45-degree line in the same way.

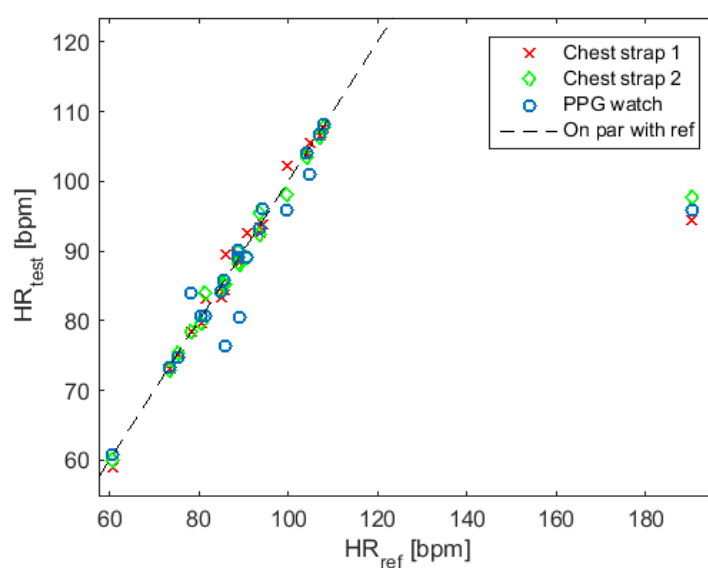
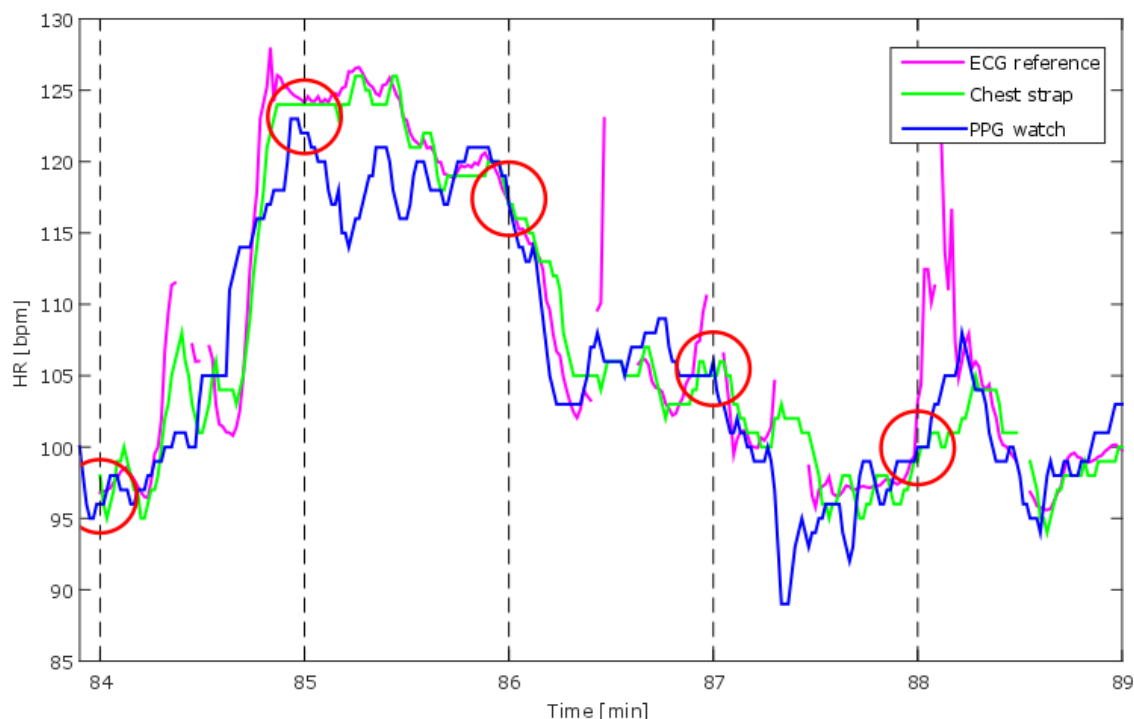


Figure 3. Segment of heart rate (HR) recordings by 3 devices: electrocardiography (ECG) reference, chest strap, and photoplethysmography (PPG) watch. The red circles represent the instants when heart rate from those devices would be collected if these were a value per minute observation. It is evident how these values do not represent the actual second by second or even the average agreement among the 3 devices.



The accuracy of the observation method should be robust (ie, repeatable and reproducible). In some validation studies, heart rate was logged manually after visually consulting the display of both ID and RD [5,7]. This method carries several limitations including human data entry errors and failure to report precisely simultaneous values from multiple devices. This method also limits the observation rate to, for instance, 1 value per minute [5,6]. Taking 1 value per minute is not the same as taking an averaged value over a minute, and both approaches fail to capitalize on the information derived from the rates of change in heart rate and heart variability and assume that participants are in a steady-state condition. Researchers should choose the observation rate (eg, 1 or 5 values per second) and averaging strategy (eg, 5- or 30-second windows) according to the use case foreseen for the heart rate monitor. Yet researchers need to be aware that taking, or averaging, 1 value every minute will hide variability [17]. This is evident in Figure 3, which illustrates that 1 single time point (red circles) is not necessarily representative of the entire minute. Consequently, for the purpose of testing accuracy, even when a mean heart rate value per minute would be sufficient, accuracy should be evaluated at the highest resolution possible.

We also observed a lack of uniformity in the statistical analyses employed in validation studies. Pearson correlations and Student *t* tests are inadequate for testing agreement [18]. This is because the Pearson correlation coefficient is not sensitive to systematic deviations from the 45-degree line, failing to reject agreement when these deviations occur. The Student *t* test is inadequate in rejecting agreement when means are equal but the 2 measures do not correlate with each other, and it can reject agreement when a very small systematic residual error shifts 1 of the means [19]. Moreover, the *t* test assesses difference, which implies

that when not rejecting the null hypothesis (ie, means are equal) it does not prove that the 2 means are equivalent. Concordance correlation coefficients should be reported instead [18,19]. Also, limits of agreement analyses should be accompanied by typical error calculations [20]. Equivalence testing should be used when the alternative hypothesis is that the outputs of 2 devices are the same [21]. In equivalence testing, the null hypothesis is that the differences between the means are outside the equivalence limits.

Finally, there are some practical considerations. The investigators should test the technology in the population of interest and in a setting appropriate for intended use. Measurements taken at rest or in the period after exercise cannot be considered to validate measurements done during exercise. Results gathered on healthy individuals with no abnormal heart rhythm are inappropriate for applications aimed at patients with cardiovascular disease where the burden of arrhythmias will be substantially higher. Additionally, due to the effect that the contact of the sensor with the skin and the environmental conditions can have on the PPG signal, information such as sensor placement, strap tightness, skin type, temperature, and possibly light intensity should be reported.

Although many studies have been published to assess the validity and usability of wrist-worn PPG-based heart rate monitoring, their methodological differences and shortcomings hamper research into their clinical utility and their introduction into health care. Such devices could make an important contribution to the future of mobile health and, in our view, should be rigorously evaluated as outlined above. For the reasons discussed in this viewpoint, we advocate standard requirements generally accepted by both the scientific

community and the device industries in order to provide a fair and consistent validation of new wearable sensor technology.

Acknowledgments

The authors would like to thank Dr Helma de Morree for reviewing the first draft of the manuscript.

Conflicts of Interest

FS and LGEC work for Royal Philips Electronics. GP is funded by Stichting voor de Technische Wetenschappen/Instituut voor Innovatie door Wetenschap en Technologie in the context of the Obstructive Sleep Apnea+ project (No 14619). JC has no conflicts of interest.

References

1. Sperlich B, Holmberg H. Wearable, yes, but able...? it is time for evidence-based marketing claims!. *Br J Sports Med* 2016 Dec 16;51(16):1240 [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/)]
2. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance).: Official Journal of the European Union; 2017 May 05. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745&from=EN> [accessed 2018-06-08]
3. US Code of Federal Regulations, Title 21: Food and Drugs, Part 820—Quality System Regulation, Subpart C—Design Controls. URL: <https://www.ecfr.gov/cgi-bin/text-idx?SID=6f0a2b9ffb8a3ac8fe2619b2381bc725&mc=true&node=se21.8.820.130&rgn=div8> [accessed 2018-06-08]
4. Quinn P. The EU commission's risky choice for a non-risk based strategy on assessment of medical devices. *Comput Law Security Rev* 2017 Jun;33(3):361-370. [doi: [10.1016/j.clsr.2017.03.019](https://doi.org/10.1016/j.clsr.2017.03.019)]
5. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol* 2016 Oct 12;2(1):104-106. [doi: [10.1001/jamacardio.2016.3340](https://doi.org/10.1001/jamacardio.2016.3340)] [Medline: [27732703](https://pubmed.ncbi.nlm.nih.gov/27732703/)]
6. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. *Ann Intern Med* 2017 Apr 18;166(8):610-612. [doi: [10.7326/L16-0353](https://doi.org/10.7326/L16-0353)] [Medline: [28395305](https://pubmed.ncbi.nlm.nih.gov/28395305/)]
7. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc* 2017 Aug;49(8):1697-1703. [doi: [10.1249/MSS.0000000000001284](https://doi.org/10.1249/MSS.0000000000001284)] [Medline: [28709155](https://pubmed.ncbi.nlm.nih.gov/28709155/)]
8. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. *J Med Internet Res* 2016 Sep 20;18(9):e253 [FREE Full text] [doi: [10.2196/jmir.6025](https://doi.org/10.2196/jmir.6025)] [Medline: [27651304](https://pubmed.ncbi.nlm.nih.gov/27651304/)]
9. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. *PLoS One* 2016;11(5):e0154420 [FREE Full text] [doi: [10.1371/journal.pone.0154420](https://doi.org/10.1371/journal.pone.0154420)] [Medline: [27232714](https://pubmed.ncbi.nlm.nih.gov/27232714/)]
10. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. *Conf Proc IEEE Eng Med Biol Soc* 2015 Aug;2015:430-433. [doi: [10.1109/EMBC.2015.7318391](https://doi.org/10.1109/EMBC.2015.7318391)] [Medline: [26736291](https://pubmed.ncbi.nlm.nih.gov/26736291/)]
11. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photoplethysmography. *Conf Proc IEEE Eng Med Biol Soc* 2014;2014:3670-3673. [doi: [10.1109/EMBC.2014.6944419](https://doi.org/10.1109/EMBC.2014.6944419)] [Medline: [25570787](https://pubmed.ncbi.nlm.nih.gov/25570787/)]
12. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med* 2017 May 24;7(2):1 [FREE Full text] [doi: [10.3390/jpm7020003](https://doi.org/10.3390/jpm7020003)] [Medline: [28538708](https://pubmed.ncbi.nlm.nih.gov/28538708/)]
13. 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/)]
14. Valenti G, Westertep K. Optical heart rate monitoring module validation study. 2013 Presented at: IEEE International Conference on Consumer Electronics; 2013; Las Vegas p. 195-196. [doi: [10.1109/ICCE.2013.6486856](https://doi.org/10.1109/ICCE.2013.6486856)]
15. ANSI/AAMI. Cardiac Monitors, Heart Rate Meters, and Alarms. Arlington: American National Standards Institute, Inc; 2002.
16. Lang M. Beyond Fitbit: a critical appraisal of optical heart rate monitoring wearables and apps, their current limitations and legal implications. *Albany Law J Sci Technol* 2017;28(1):39-72.
17. Guidance for Industry, Investigating Out-of-Specification (OOS), Test Results for Pharmaceutical Production.: Department of Health and Human Services, Center for Drug Evaluation and Research (CDER); 2006. URL: <https://www.fda.gov/downloads/drugs/guidances/ucm070287.pdf> [accessed 2018-06-08] [WebCite Cache ID 701iuEKAU]
18. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int J Cardiol* 2013 Jun 05;166(1):15-29. [doi: [10.1016/j.ijcard.2012.03.119](https://doi.org/10.1016/j.ijcard.2012.03.119)] [Medline: [22809539](https://pubmed.ncbi.nlm.nih.gov/22809539/)]

19. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989 Mar;45(1):255-268. [Medline: [2720055](#)]
20. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000 Jul;30(1):1-15. [Medline: [10907753](#)]
21. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc Psychol Personal Sci* 2017 May;8(4):355-362 [[FREE Full text](#)] [doi: [10.1177/1948550617697177](#)] [Medline: [28736600](#)]

Abbreviations

ECG: electrocardiography
ID: investigational device
PPG: photoplethysmography
RD: reference device

Edited by G Eysenbach; submitted 13.02.18; peer-reviewed by M Lang, P Wark; comments to author 22.03.18; revised version received 16.05.18; accepted 29.05.18; published 02.07.18

Please cite as:

Sartor F, Papini G, Cox LGE, Cleland J

Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations

J Med Internet Res 2018;20(7):e10108

URL: <http://www.jmir.org/2018/7/e10108/>

doi: [10.2196/10108](#)

PMID: [29967000](#)

©Francesco Sartor, Gabriele Papini, Lieke Gertruda Elisabeth Cox, John Cleland. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.07.2018. 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 the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.