Health & Fitness Journal of Canada

Volume 16 **March 30, 2023** **Number 1**

EXPERT OPINION

Establishing Validity and Reliability of HRV Smartphone Applications: What Have We Really Been Examining?

Andrew S. Perrotta[1,2,*] and Elizabeth A. Gnatiuk[3]

1 Department of Kinesiology, Faculty of Human Kinetics, University of Windsor, Windsor, Ontario, Canada, Canada N9B 3P4
2 Centre for Human Performance and Health, University of Windsor, Windsor, Ontario, Canada, N9B 3P4
3 Canadian Sport Institute Pacific, Victoria, British Columbia, Canada, V9E 2C5
Corresponding Author: perrotta@uwindsor.ca

Abstract

Background: Heart rate variability (HRV) has become a regular metric for practitioners assessing the dose-response to exercise. The utility of smartphone applications capable of examining HRV using wearable technology has made this form of cardiac assessment more efficient and effective outside of laboratory settings. **Purpose**: This opinion piece will discuss the various forms of validity and reliability within research, and their applications towards establishing the accuracy and consistency of HRV smartphone applications. **Conclusions:** The unremitting creation of new HRV smartphone applications is cause for the continued examination of their validity and reliability. Understanding that the definition of validity has a broad meaning, requires researchers to state the exact form of validity they intend to examine. Furthermore, all forms of validity must be meticulously examined and viewed as equivalent when establishing the accuracy of a smartphone application. When assessing the reliability of an HRV application, researchers should focus on determining the consistency of its error in relation to the gold standard using repeated trials. **Health & Fitness Journal of Canada 2023;16(1):3-9**.
https://doi.org/10.14288/hfjc.v16i1.827

Keywords: Heart Rate Variability, Wearable Technology, Cardiovascular Function, Sport Science

Introduction

The clinical relevance of heart rate variability (HRV), as defined through the modulation of beat-to-beat intervals, to reflect cardiac autonomic activity (Hon & Lee, 1963), has supported clinicians in the monitoring of patients with cardiovascular disease (Goldenberg et al., 2019; Sessa et al., 2018). The utility of HRV outside of a clinical setting has significantly expanded since the 21st century. To date, there are numerous reviews discussing HRV and how to examine changes in vagal activity as they equate to alterations in human performance (Aubert, Seps, & Beckers, 2003; Buchet, 2014; Stanley, Peake, & Buchheit, 2013). HRV was primarily shown to be valuable for monitoring the dose-response to exercise in individualized sporting events (Plews, Laursen, Kilding, & Buchheit, 2012; Plews, Laursen, Stanley, Kilding, & Buchheit, 2013; Plews, Laursen, Kilding, & Buchheit, 2014). However, continuing research has demonstrated its efficacy for use in team sports, such as soccer (Flatt & Esco, 2015), rugby

(Nakamura et al., 2017a), basketball (Nakamura et al., 2017b), American football (Flatt et al., 2020), and field hockey (Perrotta, Koehle, White, Taunton, & Warburton, 2019; Perrotta, 2020).

With the ever-increasing popularity of HRV, wearable technology has rapidly evolved to provide practitioners with a user-friendly experience. The introduction of smartphone applications that communicate via Bluetooth® to heart rate monitors has led to the unremitting growth of proprietary technology and software. Expert statements, such as the Network of Physical Activity Assessment (INTERLIVE), a joint European initiative involving universities and industrial partners, continue to develop recommendations for assessing the validity and reliability of wearable technology (Mühlen et al., 2021). These recommendations can support systematic reviews and meta-analyses that investigate the accuracy of smartphone applications (Dobbs et al., 2019). Still, continuing inquiry into the validity of identical smartphone applications display diverse results (Moya-Ramon, Mateo-March, Peña-González, Zabala, & Javaloyes, 2022; Perrotta, Jeklin, Hives, Meanwell, & Warburton, 2017). These differences are a consequence of the unique methodology, as well as different interpretations of what validity and reliability imply.

The current methodology for establishing the validity and reliability of HRV equipment has substantially deviated from the recommendations and standards set forth by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Variability, 1996).

This opinion piece will discuss the different forms of validity within research, and how reliability can be determined when examining HRV smartphone applications.

What is the Definition of Validity?

The term "validity" in this context can be defined as the level of accuracy an HRV smartphone application provides its user. When examining the validity of an HRV application, researchers must choose the type of validation they are interested in and clearly define the term for the reader. Only then can acceptable transparency be presented to support the observations and allow for the replication of future studies.

The research methodology that is designed to examine the validity of an HRV smartphone application can have four different forms (Hall & Getchell, 2014).

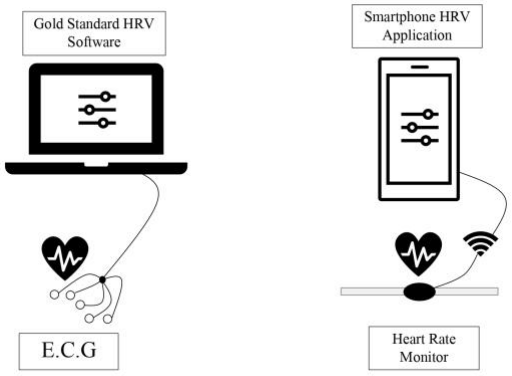
I. Criterion Validity
II. Logical (Face) Validity
III. Content Validity
IV. Construct Validity

Criterion Validity

Criterion validity is often labelled as the 'gold standard" for assessing the accuracy of a new instrument. A sub-set of this form is known as "concurrent validity". This type of validation compares the accuracy of a new instrument to a recognized criterion instrument (Thomas, Martin, Etnier, & Silverman, 2022). Many studies examining the validity of new HRV applications have focused on this type of validation. When examining the concurrent validity of an HRV application, it must involve the simultaneous collection of R-R intervals of a single patient using a criterion form (i.e., electrocardiogram) and a secondary device (i.e., heart rate monitor). This form of validation involves R-R intervals being exported via Bluetooth® from the wearable device directly to the smartphone application for analysis using its

proprietary software. R-R intervals collected via ECG are exported to a validated computer software program, such as Kubios HRV® (Tarvainen, Niskanen, Lipponen, Ranta-aho, & Karjalainen, 2014) for analysis. It is important to recognize, that criterion validity does not focus solely on the legitimacy of the HRV smartphone application itself. When examining criterion validity, researchers must be cognisant that both pieces of equipment (i.e., ECG and heart rate monitors), even when possessing identical sampling frequency (1000 Hz), have innate error, and provide unique R-R interval recordings (Giles, Draper, & Neil, 2016). As such, comparison between HRV values that are derived from a smartphone application and a computer software program will inherently be different due to differences in the hardware, not the smartphone application itself (Figure 1). Current literature has shown strong agreement between commercially available heart rate monitors and ECG for recording beat-to-beat intervals (Schaffarczyk, Rogers, Reer, & Gronwald, 2022).

Figure 1: Technique for establishing criterion validity.



Logical, Content, and Construct Validity

The examination of logical, content and construct validity, can be viewed as a comprehensive, yet concentrated analysis of an HRV smartphone application. Logical (face) validity examines the degree to which the application appears to measure HRV. Content validity examines how well an application covers all applicable aspects of HRV measurement, while construct validity examines the extent to which the application accurately assesses HRV (Thomas et al., 2022).

By nature, HRV smartphone applications are proprietary software algorithms that are unique to each company. Therefore, the factual validation of a smartphone application is the examination of its software in comparison to the gold standard software used to derive HRV. Recommendations for examining the validity of innovative HRV software have advocated for the use of simulated R-R interval recordings, to negate differences in R-R intervals when recorded using two devices (Variability, 1996)

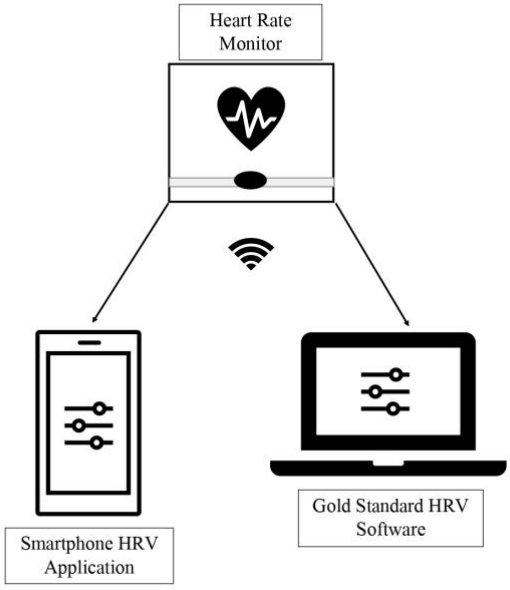
The important distinction between these forms of validity and criterion validity involves the collection of R-R intervals using a singular instrument (i.e., heart rate monitor). When R-R intervals are recorded using a singular device, researchers can eliminate equipment error and focus solely on the error derived from the smartphone application itself (Figure 2). Few investigations have specifically examined these forms of validity within an HRV application (Perrotta et al., 2017). HRV indices are derived using standardized mathematical equations. This would suggest any differences in HRV values between an application and a computer software program are likely due to artifact correction techniques. This is an

important aspect that needs to be clearly discussed and made transparent in the methodology of a research project. The use of artifact correction techniques are often left unstandardized. Various reviews have discussed artifact correction techniques as a confounding factor, and the need to standardize techniques to enhance the validity and reliability of HRV monitoring (Alcantara et al., 2020; Peltola, 2012). Smartphone applications often involve proprietary artifact correction algorithms that prevent researchers from being transparent in their methodologies.

Statistical Approaches for Establishing Validity

Validation of a smartphone application must be established using appropriate statistical analysis. It has been recommended that the agreement between a novel instrument and the accepted gold standard be investigated rather than solely examining the association between the two (Bland & Altman, 1986; Giavarina, 2015). Utilizing a Bland-Altman analysis allows for the identification of the mean bias, as well as the rigidity of the limits of agreement. It has been suggested that 95% of the differences between the new instrument and the gold standard should reside within ±2.0 standard deviations of the mean difference to be considered an acceptable alternative (Bland & Altman, 1986). Establishing validity can be calculated over multiple trials. A repeated measures analysis can be implemented to identify significant differences (*p* value) between instruments. Depending on the number of assessments, a one-way ANOVA or a *t* test can be conducted. Additionally, the magnitude of difference between instruments can be determined by calculating the standardized mean differences and presented using Cohen's *d* (Cohen, 1988).

Figure 2: Technique for establishing content, logical, and construct validity.



How Do We Establish the Reliability of an HRV Application?

The term "reliability" can be defined as the repeatability of a test to produce similar results (Thomas et al., 2022). HRV smartphone applications continue to be examined for their reliability (Moya-Ramon et al., 2022). Reliability experiments typically utilize a test re-rest trial, comparing raw HRV values taken in short succession to establish the magnitude of variation between trials. However, cardiac autonomic activity is in continuous fluctuation in response to biological functioning. Any comparison in repeated measures of HRV values is a reflection of the magnitude of change in biological functioning, rather than the reliability of the instrument or software to record and analyze R-R intervals. Comparing repeated HRV values is not practical nor appropriate for examining the

reliability of a smartphone application to derive accurate values. Instead, researchers should focus on examining the magnitude of error the smartphone application demonstrates between trials. This error can be expressed using either the residuals or the mean absolute error when calculating the difference between the smartphone application and the gold standard software.

Statistical Approaches for Establishing Reliability

Once the error is calculated over multiple trials, researchers can utilize different approaches for examining the reliability of smartphone applications. One approach is a repeated measures analysis. This technique can identify significant differences (*p* value) in the application error between trials. This assessment can utilize a one-way ANOVA or a *t* test depending on the number of assessments conducted. Additionally, the coefficient of variation (CV %) can be determined using this technique. Researchers can also quantify the magnitude of variation in error between trials using Cohen's *d*.

A secondary assessment, such as the standard error of measurement (SEM) can be conducted using a repeated measures analysis. However, it is important to recognise that the correlation coefficient (*r*) used in the SEM equation, must be derived from the association in the error between multiple tests, rather than the raw HRV values.

Conclusions

The persistent manufacturing of HRV smartphone applications is cause for continued examination of their validity and reliability. Understanding that the definition of validity has a broad meaning, requires researchers to state the exact form of validity they intended to examine. Furthermore, all forms of validity must be meticulously examined and viewed as equivalent when establishing the accuracy of a smartphone application. When assessing the reliability of an HRV application, researchers should focus on determining the consistency of its error in relation to the gold standard using repeated trials.

Acknowledgements

The authors would like to endearingly thank Dr. Darren E.R. Warburton for his mentorship and support.

Authors' Qualifications

The authors' qualifications are as follows: Andrew S. Perrotta Ph.D., M.Kin., B.A.Sc. (HON), CEP, HPS, CSCS; Elizabeth A. Gnatiuk M.Kin, BSc., CEP, HPS, CSCS, ISAK Level 2.

References

Alcantara, J. M., Plaza-Florido, A., Amaro-Gahete, F. J., Acosta, F. M., Migueles, J. H., Molina-Garcia, P., … Martinez-Tellez, B. (2020). Impact of using different levels of threshold-based artefact correction on the quantification of heart rate variability in three independent human cohorts. *Journal of Clinical Medicine*, *9*(2), 325. doi:10.3390/jcm9020325

Aubert, A. E., Seps, B., & Beckers, F. (2003). Heart rate variability in athletes. *Sports Medicine*, *33*(12), 889–919. doi:10.2165/00007256-200333120-00003

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310. doi:10.1016/s0140-6736(86)90837-8

Buchheit, M. (2014). Monitoring training status with HR measures: Do all roads lead to Rome? *Frontiers in Physiology*, *5*. doi:10.3389/fphys.2014.00073

Camm, A. J., Malik, M., Bigger, J. T., Breithardt, G., Cerutti, S., Cohen, R. J., ... & Singer, D. H. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European

Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation, 93(5), 1043-1065.

Cohen, J. (1988). *Statistical Power Analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates Inc.

Dobbs, W. C., Fedewa, M. V., MacDonald, H. V., Holmes, C. J., Cicone, Z. S., Plews, D. J., & Esco, M. R. (2019). The accuracy of acquiring heart rate variability from portable devices: A systematic review and meta-analysis. *Sports Medicine*, *49*(3), 417–435. doi:10.1007/s40279-019-01061-5

Flatt, A. A., & Esco, M. R. (2015). Smartphone-derived heart-rate variability and training load in a women's soccer team. *International Journal of Sports Physiology and Performance*, *10*(8), 994–1000. doi:10.1123/ijspp.2014-0556

Flatt, A. A., Allen, J., Bragg, A., Keith, C., Earley, R., & Esco, M. R. (2020). Heart rate variability in college football players throughout preseason camp in the heat. *International Journal of Sports Medicine*, *41*(09), 589–595. doi:10.1055/a-1145-3754

Giavarina, D. (2015). Understanding bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. doi:10.11613/bm.2015.015

Giles, D., Draper, N., & Neil, W. (2016). Validity of the polar V800 heart rate monitor to measure RR intervals at rest. *European Journal of Applied Physiology*, *116*(3), 563–571. doi:10.1007/s00421-015-3303-9

Goldenberg, I., Goldkorn, R., Shlomo, N., Einhorn, M., Levitan, J., Kuperstein, R., … Johnson, B. (2019). Heart rate variability for risk assessment of myocardial ischemia in patients without known coronary artery disease: The HRV-DETECT (Heart Rate Variability for the Detection of Myocardial Ischemia) Study. *Journal of the American Heart Association*, *8*(24). doi:10.1161/jaha.119.014540

Hall, S., & Getchell, N. (2014). *Research methods in kinesiology and the Health Sciences* (1st ed.). Wolters Kluwer Health.

Hon, E. H., & Lee, S. T. (1963). Electronic evaluation of the fetal heart rate. *American Journal of Obstetrics and Gynecology*, *87*, 814–826. doi:10.1016/s0002-9378(16)35841-0

Moya-Ramon, M., Mateo-March, M., Peña-González, I., Zabala, M., & Javaloyes, A. (2022). Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes. *Computer Methods and Programs in Biomedicine*, *217*, 106696. doi:10.1016/j.cmpb.2022.106696

Mühlen, J. M., Stang, J., Lykke Skovgaard, E., Judice, P. B., Molina-Garcia, P., Johnston, W., … Schumann, M. (2021). Recommendations for determining the validity of Consumer Wearable Heart Rate Devices: Expert statement and checklist of the INTERLIVE network. *British Journal of Sports Medicine*, *55*(14), 767–779. doi:10.1136/bjsports-2020-103148

Nakamura, F. Y., Pereira, L. A., Esco, M. R., Flatt, A. A., Moraes, J. E., Cal Abad, C. C., & Loturco, I. (2017a). Intraday and Interday reliability of ultra-short-term heart rate variability in rugby union players. *Journal of Strength and Conditioning Research*, *31*(2), 548–551. doi:10.1519/jsc.0000000000001514

Nakamura, F. Y., Pereira, L. A., Cal Abad, C. C., Cruz, I. F., Flatt, A. A., Esco, M. R., & Loturco, I. (2017b). Adequacy of the ultra-short-term HRV to assess adaptive processes in youth female basketball players. *Journal of Human Kinetics*, *56*(1), 73–80. doi:10.1515/hukin-2017-0024

Peltola, M. A. (2012). Role of editing of R–R intervals in the analysis of heart rate variability. *Frontiers in Physiology*, *3*, 148. doi:10.3389/fphys.2012.00148

Perrotta, A. S., & Warburton, D. E. R. (2020). Alterations in cardiac vagal modulation-to-vagal tone ratio in response to accumulated exercise stress in intermittent team sport. *Biomedical Human Kinetics*, *12*(1), 197–203. doi:10.2478/bhk-2020-0025

Perrotta, A. S., Jeklin, A. T., Hives, B. A., Meanwell, L. E., & Warburton, D. E. R. (2017). Validity of the elite HRV smartphone application for examining heart rate variability in a field-based setting. *Journal of Strength and Conditioning Research*, *31*(8), 2296–2302. doi:10.1519/jsc.0000000000001841

Perrotta, A. S., Koehle, M. S., White, M. D., Taunton, J. E., & Warburton, D. E. R. (2019). Consecutive non-training days over a weekend for assessing cardiac parasympathetic variation in response to accumulated exercise stress. *European Journal of Sport Science*, 1–11. doi:10.1080/17461391.2019.1688397

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2012). Heart rate variability in elite

triathletes, is variation in variability the key to effective training? A case comparison. *European Journal of Applied Physiology*, *112*(11), 3729–3741. doi:10.1007/s00421-012-2354-4

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. *Sports Medicine*, *43*(9), 773–781. doi:10.1007/s40279-013-0071-8

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2014). Heart-rate variability and training-intensity distribution in elite rowers. *International Journal of Sports Physiology and Performance*, *9*(6), 1026–1032. doi:10.1123/ijspp.2013-0497

Schaffarczyk, M., Rogers, B., Reer, R., & Gronwald, T. (2022). Validity of the Polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. *Sensors*, *22*(17), 6536. doi:10.3390/s22176536

Sessa, F., Anna, V., Messina, G., Cibelli, G., Monda, V., Marsala, G., … Salerno, M. (2018). Heart rate variability as predictive factor for sudden cardiac death. *Aging*, *10*(2), 166–177. doi:10.18632/aging.101386

Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: Implications for training prescription. *Sports Medicine*, *43*(12), 1259–1277. doi:10.1007/s40279-013-0083-4

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. doi:10.1016/j.cmpb.2013.07.024

Thomas, J. R., Martin, P. E., Etnier, J. L., & Silverman, S. J. (2022). *Research methods in physical activity*. Champaign, IL: Human Kinetics, Inc.

Variability, H. R. (1996). Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation*, *93*, 1043-1065.