TOPICS IN EXERCISE SCIENCE AND KINESIOLOGY

Process of Science

Data Analysis Processes and Techniques for Validation of Wearable Technology: An Example

BRYSON CARRIER and JAMES W. NAVALTA

Department of Kinesiology and Nutrition Sciences, University of Nevada, Las Vegas, Las Vegas, NV, USA

ABSTRACT

Topics in Exercise Science and Kinesiology Volume 3: Issue 1, Article 10, 2022. With wearable technology growing in popularity and sophistication, there remains a need to determine the validity of these devices by independent observers. Validation studies of wearable technology can involve large amounts of data, with data preparation techniques that are not always clearly established. This can make attempts to reproduce the results difficult and does not allow researchers to gain guidance in how to perform their own analyses if they wanted to perform a similar study. Therefore, this paper details the process that was utilized to prepare and analyze the accuracy of several heart rate monitors during mountain biking and can be used as a possible guide to researchers looking to perform similar analyses. We also detail the software used and discuss possible alternatives.

KEY WORDS: Fitness tracker, activity monitor, biosensor, data processing, sport science

INTRODUCTION

Wearable technology continues to grow in popularity and sophistication, securing the top slot in a worldwide survey of fitness trends 6 of the last 7 years *(1-7)*. With this popularity and continued evolution of the device features, there remains a need to understand the validity of these devices *(8)*. Wearable technology can now estimate or measure a variety of factors, including heart rate, VO2max, lactate threshold, steps, energy expenditure, and many other aspects. Validating these devices through independent observers is of increasing importance, as more people use them for recreation, professional, and even research purposes. When validating these devices, there is not widespread agreement on the process, tests, or thresholds to determine validity. Data collection and analysis in wearable technology can be a difficult task, as it often requires analysis of large datasets. As was the case for our lab group as we sought to validate several heart rate monitors during mountain biking. There have been analytical techniques suggested by multiple authors in terms of tests to use *(8, 9)*, however, the process of preparing the data and analyzing it is a task that many researchers could use additional guidance to perform properly. The current paper is meant to serve as an example of a pathway other researchers could take when analyzing data for their own projects. This is not meant to be a paper of suggested best-practices of data analysis in wearable technology. Detailing this type of analysis will become more important as the amount of data that wearable technology provides continually increases. Up to this point, many studies utilizing wearable technology did not look at second-by-second data, thus having a lesser amount of data to wrangle, requiring fewer data analysis techniques to properly prepare the data for statistical analysis *(10, 11)*. As the consumer technology association (CTA) recommends that wearable technology be analyzed second-to-second, when possible *(12)*, this paper can give insights into the possible ways to perform this type of analysis in the future.

Additionally, it has been our experience that there is reluctance among authors, reviewers, and editors to have such a detailed process in the validation literature. While this may simply be our experience with the journals we have interacted with, we think others will have had similar experiences. Therefore, this article will provide an example, detailing the process of data collection, preparation, and analysis used to validate the heart rate monitors in the current investigation.

METHODS

Twenty apparently healthy participants (10 male, 10 female, age = 26.3 ± 6.6 years, height = 171.8 ± 8.0cm, mass = 73.9 ± 19.0kg, reported as mean ± SD) completed two self-paced mountain biking trials (3.22km) while wearing six devices (5 test devices and 1 criterion). There were 17 individuals who self-identified as having low MTB experience, three with moderate MTB experience, and none that reported having high MTB experience. The device information for all devices used can be found in table 1. The Polar H7 Heart Rate Monitor which utilizes ECG technology to determine HR was used as the criterion device. This device has previously been found to have high agreement with ambulatory ECG devices *(13-17)*. However, due to a technical issue resulting in a failure to collect data of the criterion device for the final four subjects, only data from 16 participants were included in the analysis.

Table 1. Device and Company Information.

Brand	Device	Company Information
Garmin	fēnix® 5	Garmin Ltd., Schaffhausen, Switzerland
Jabra	Elite Sport Earbuds	Jabra, Copenhagen, Denmark
Suunto	Spartan Sport Watch + Chest HRM	Suunto Oy, Vantaa, Finland
Scosche	Rhythm+	Scosche Industries Inc., Oxnard, CA, USA
Polar	H7 Heart Rate Monitor	Polar Electro Inc., Woodbury, NY, USA
Polar	A360 Fitness Tracker	Polar Electro Inc., Woodbury, NY, USA

Company information of each device used in the current study. Polar H7 Heart Rate monitor used as criterion device.

DATA ANALYSIS AND RESULTS

Device Set-up and Data Extraction

The devices used in the current investigation were all updated the night before each test. The devices that were compatible, were connected to a third-party app, PerformTek Data Collector (Valancell Inc, Raleigh, North Carolina, USA), that was used to compile all the data for convenience in analysis. The devices that were not able to connect to this app were downloaded separately as an excel or CSV file. The exception to this was the Garmin fenix 5, which outputs as a GPX file. Therefore, custom Python code was used to convert these files to CSV. After the data was extracted from each device, the files were converted to CSV format (if needed) and joined by the date and time stamp via custom Python code in Homebrew (Software Freedom Conservancy, Brooklyn, NY, USA). All devices produced results in a second-by-second format, and values were expressed as beats per minute (bpm).

Data Trimming Procedure

The data was then trimmed at the beginning and end to account for varying start and end times of the devices due to each device being started and stopped manually by the researchers. The data was trimmed until all devices were recording. There was an average of 26 seconds removed from each end. After the data was trimmed, a quality assessment of the criterion device data was performed, and where null data, "0" values, or abnormal data in the criterion device was found, the data at that time was removed from all devices. Finally, non-physiological data points were removed from any device (bpm>220). There was a total of 35,774 lines of data after data processing was completed. See table 2 for a breakdown of the data removal steps.

Table 2. Data Processing and Removal Steps.

Data Points from Original
37,674
Data Points After Trimming Ends
36,034
Total Data Points Removed From Trimming Ends
1,640
Total Time Removed from Trimming Ends in Entire Dataset (min)
27.33
Avg Time Removed From Each Trial (sec)
51.25
Avg Time Removed from Each End of Trial (sec)
25.63
Total Data Points Removed Due to Non-Physiological Values (>220 bpm)
13 (all from Rhythm+ Device)
Data Points After Removing 0's and Other Abnormal Data from Criterion Device Data
35,774 (260 lines removed)

Documentation of the data processing and data removal steps taken.

Validation Measures

Validation measures were obtained by comparing the results of the combined trials of the test device to the criterion measure at each second. The data was then stratified into five HR phases based on the mean age of the participants and validity measures determined for each stratified dataset. Validity was determined for each analysis via multiple statistical tests: 1. Error analysis, mean absolute percentage errors (MAPE), mean absolute error (MAE), and mean error (ME); 2. Correlation analysis, Lin's concordance correlation coefficient (CCC) and Pearson's correlation coefficient (r); and 3. Equivalence Testing (two one-way t-tests [TOST test]). Thresholds++ for validity were predetermined, based on previous publications *(8, 18, 19)*. A MAPE of <10% and a CCC value of >0.7 would result in a valid classification for that device. While TOST tests were performed for each device, the results were not considered in the validation threshold criteria as appropriate thresholds for TOST testing have yet to be established for wearable technology. A device had to satisfy thresholds for both statistical tests to be considered valid. All statistical analyses were performed using Google Sheets (Google LLC, Mountain View CA, USA), SPSS (Version 24.0, International Business Machines Corp. [IBM], Armonk, NY, USA), and jamovi (The jamovi project [2021]. jamovi Version 1.6 [Computer Software]. Retrieved from https://www.jamovi.org). Any values registered as a HR of "0" were not factored into the averages.

The time per trial was determined by calculating the time between the first timestamp of a trial, and the last timestamp. Means and standard deviations were calculated for the data, and a one-tailed, paired t-test was performed on the mean trial times, and the coefficient of determination (r^2) was calculated for each device. Demographic data was also collected and means and standard deviations were calculated.

Data Characterization

We have also included the results for the data characterization of the devices for the convenience of the reader (see Table 3). The devices that had the greatest data availability, (measured as a percentage of available data points compared to the criterion) in descending order, were the Suunto (99.95%), Rhythm+ (97.17%), fenix 5 (96.44%), Polar A360 (92.66%), and finally the Jabra (22.27%).

Table 3. Data Characterization.

	Polar H7 Chest HRM	Suunto Chest HRM	Rhythm+ HR Monitor	fēnix 5x Watch	Polar A360 Watch	Jabra Earbuds
Total 0's	0	0	999	0	0	0
Total Null Values	0	19	13	38	42	27807
Total "-"	0	0	0	1235	2584	0
Summed 0, Null, and "-" Values	0	19	1012	1273	2626	27807
Total Data Points in Dataset	35774	35755	34762	34501	33148	7967
Data Availability (Percent of Criterion)		99.95%	97.17%	96.44%	92.66%	22.27%

Breakdown of the total number of non-normal values, total data points, and data availability by device (n=16).

DISCUSSION

As wearable technology continues to grow in popularity and sophistication, researchers will have to use more sophisticated data analysis techniques to deal with the influx of data provided by the wearable devices. This can be done with the use of several different software solutions currently available. For those that know how to write code, Python (Python Software Foundation, https://www.python.org/) and R (RStudio PBC, https://www.rstudio.com/) are common data analysis languages used for this type of analysis. For the current investigation, we were fortunate enough to have a researcher that was able to assist in the difficult parts of the data preparation and analysis by writing custom Python code. However, not everyone will have the ability to perform this type of analysis. Difficult data preparation is still possible without needing to write code. Software such as Tableau Prep Builder (Tableau Software Inc, Seattle, WA, USA) or KNIME (KNIME AG, Zürich, Zurich, Switzerland) are good alternatives for preparing data without needing to know how to write code. Tableau Prep is a paid service that accompanies the Tableau business intelligence software, and KNIME is an open-source software. These point-and-click options can perform more complex tasks needed to prepare the data, such as joins, filters, compilations, and other transformations that would be very difficult given traditional statistical software programs. They are designed for large-scale data analysis, and therefore are suitable for large wearable technology datasets. However, they can be resource intensive and may require more robust computers to complete the analysis. For researchers looking to prepare and analyze large datasets that often accompany wearable technology, they are a viable option. When researchers are determining the best course of action to prepare and analyze their data, the processes detailed in this paper can guide their decision making.

Wearable technology needs to be properly evaluated to determine the accuracy and validity by independent parties. This analysis may require sophisticated data analysis techniques to deal with large datasets. Software and coding languages can assist in this process, and the specific process undertaken by our research group to validate several heart rate monitors during mountain biking can be found in the current manuscript. Those seeking to perform similar analyses may find this paper useful to guide them in possible routes for completing their own data analyses in the future.

REFERENCES

1. Thompson, W. R. Worldwide survey of fitness trends for 2016. *ACSM's Health & Fitness Journal* **2015**, *19*, 9-18.

2. Thompson, W. R. WORLDWIDE SURVEY OF FITNESS TRENDS FOR 2017. *ACSM's health & fitness journal* **2016**, *20*, 8-17.

3. Thompson, W. R. WORLDWIDE SURVEY OF FITNESS TRENDS FOR 2018: The CREP Edition. *ACSM's health & fitness journal* **2017**, *21*, 10-19.

4. Thompson, W. R. WORLDWIDE SURVEY OF FITNESS TRENDS FOR 2019. *ACSM's health & fitness journal* **2018**, *22*, 10-17.

5. Thompson, W. R. WORLDWIDE SURVEY OF FITNESS TRENDS FOR 2020. *ACSM's health & fitness journal* **2019**, *23*, 10-18.

6. Thompson, W. R. Worldwide Survey of Fitness Trends for 2021. *ACSM's health & fitness journal* **2021**, *25*, 10-19.

7. Thompson, W. R. Worldwide survey of fitness trends for 2022. *ACSM's Health & Fitness Journal* **2022**, *26*, 11-20.

8. Carrier, B.; Barrios, B.; Jolley, B. D.; Navalta, J. W. Validity and Reliability of Physiological Data in Applied Settings Measured by Wearable Technology: A Rapid Systematic Review. *Technologies* **2020**, *8*, 70.

9. Welk, G. J.; Bai, Y.; Lee, J.; Godino, J.; Saint-Maurice, P. F.; Carr, L. Standardizing analytic methods and reporting in activity monitor validation studies. *Med. Sci. Sports Exerc.* **2019**, *51*, 1767.

10. Borges, N. R.; Driller, M. W. Wearable lactate threshold predicting device is valid and reliable in runners. *Journal of strength and conditioning research* **2016**, *30*, 2212-2218.

11. Carrier, B.; Creer, A.; Williams, L. R.; Holmes, T. M.; Jolley, B. D.; Dahl, S.; Weber, E.; Standifird, T. Validation of garmin fenix 3 HR fitness tracker biomechanics and metabolics (VO2max). *Journal for the Measurement of Physical Behaviour* **2020**, *3*, 331-337.

12. Consumer Technology Association Physical Activity Monitoring for Heart Rate, ANSI/CTA-2065. **2018**.

13. Hernández-Vicente, A.; Hernando, D.; Marín-Puyalto, J.; Vicente-Rodríguez, G.; Garatachea, N.; Pueyo, E.; Bailón, R. Validity of the Polar H7 Heart Rate Sensor for Heart Rate Variability Analysis during Exercise in Different Age, Body Composition and Fitness Level Groups. *Sensors* **2021**, *21*, 902.

14. Kingsley, M.; Lewis, M. J.; Marson, R. E. Comparison of polar 810 s and an ambulatory ECG system for RR interval measurement during progressive exercise. *Int. J. Sports Med.* **2005**, *26*, 39-44.

15. Weippert, M.; Kumar, M.; Kreuzfeld, S.; Arndt, D.; Rieger, A.; Stoll, R. Comparison of three mobile devices for measuring R–R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *Eur. J. Appl. Physiol.* **2010**, *109*, 779-786.

16. Dobbs, W. C.; Fedewa, M. V.; MacDonald, H. V.; Holmes, C. J.; Cicone, Z. S.; Plews, D. J.; Esco, M. R. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. *Sports Medicine* **2019**, *49*, 417-435.

17. Pasadyn, S. R.; Soudan, M.; Gillinov, M.; Houghtaling, P.; Phelan, D.; Gillinov, N.; Bittel, B.; Desai, M. Y. Accuracy of commercially available heart rate monitors in athletes: a prospective study. *Cardiovascular diagnosis and therapy* **2019**, *9*, 379.

18. Navalta, J. W.; Montes, J.; Bodell, N. G.; Salatto, R. W.; Manning, J. W.; DeBeliso, M. Concurrent heart rate validity of wearable technology devices during trail running. *Plos one* **2020**, *15*, e0238569.

19. Carrier, B.; Creer, A.; Williams, L. R.; Holmes, T. M.; Jolley, B. D.; Dahl, S.; Weber, E.; Standifird, T. Validation of Garmin Fenix 3 HR Fitness Tracker Biomechanics and Metabolics (VO2max). *Journal for the Measurement of Physical Behaviour* **2020**, *3*, 331-337.

Data Analysis Processes and Techniques for Validation of Wearable Technology: An Example

BRYSON CARRIER and JAMES W. NAVALTA
Department of Kinesiology and Nutrition Sciences, University of Nevada, Las Vegas, Las Vegas, NV, USA

SUMMARY
Wearable technology is constantly growing in popularity and sophistication. Tools and techniques to perform data analysis on the ever-increasing amount of data should be of interest to researchers. This paper detailed the process taken by our research team to prepare and analyze the data of several heart rate monitors during mountain biking to independently validate the accuracy of each monitor. Twenty apparently healthy individuals completed two mountain-bike trials at a self-selected pace while wearing six heart rate monitors. The heart rate monitors tested were the Garmin fēnix® 5, Jabra Elite Sport Earbuds, Suunto Spartan Sport Watch + Chest HRM, Scosche Rhythm+, Polar H7 Heart Rate Monitor, and the Polar A360.

The data were recorded through a third-party app used (PerformTek) which output the HR in second-by-second format. The devices that were not able to connect to the PerformTek app were exported individually to a CSV, with the exception of the Garmin fenix 5, which had to be exported as a GPX file and converted to CSV via custom Python code. The data was trimmed on each end to account for different start and stop times, as each device was manually started. Null values, "0" values, and other abnormal values in the criterion device were also excluded from the analysis.

Validation measures to determine the validity of the device were measured through three categories of tests. 1. Error analysis, 2. Correlation analysis, and 3. Equivalence analysis. Pre-determined thresholds were established to determine the validity of the device. A mean absolute percentage error of <10% and a Lin's concordance correlation coefficient of >0.7 classified that device as valid.

The devices that had the greatest data availability, (measured as a percentage of available data points compared to the criterion) in descending order, were the Suunto (99.95%), Rhythm+ (97.17%), fenix 5 (96.44%), Polar A360 (92.66%), and finally the Jabra (22.27%).

The tools utilized for this analysis were several statistical software programs, including custom code written by one of the researchers. Those looking to perform research with wearable technology who may need to deal with large datasets could use data preparation software like Tableau Prep and KNIME, or coding languages such as R and Python.