A Statistical Comparison of Heart Rate Variability Measurements Between Devices: Chest Strap Versus Finger Probe

Mehri Bagheri-Mohamadi-Pour[1], Rodney Sparapani[2], Jun Zhang[1], Jacquelyn Kulinski[3]

[1] Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[2] Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA
[3] Division of Cardiovascular Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

Abstract

Heart rate variability (HRV) between R-R intervals results from complex interactions between respiratory activity and autonomic cardiovascular control. Measurements of HRV are strong predictors of cardiac morbidity and mortality. A finger probe is more convenient than a chest strap device so we compared the HRV measurements between the CorSense finger probe and the Polar H7 chest strap. This data was collected in our clinical study (NCT04121741) of 33 subjects (mean age 69 with standard deviation 10, 52% female) with coronary heart disease. We captured HRV in both the time-domain (root mean square of successive RR interval differences (rMSSD), standard deviation of NN intervals (SDNN), percentage of successive RR intervals that differ by more than 50 ms (PNN50)) and the frequency-domain (low-frequency power (LF Power) and high-frequency power (HF Power)). Regression analyses and the Wilcoxon signed-rank-test (WSRT) showed a closer correlation between devices for SDNN and LF Power. However, none of these metrics met the satisfactory agreement criteria. The clinical study had singing and non-singing control sessions. Due to possible discrepancies while singing, we only included non-singing control sessions.

1. Introduction

Our study evaluates the reliability and performance of the CorSense finger probe versus the Polar H7 chest strap for measuring HRV in older adults with cardiovascular disease. HRV captures variations between consecutive heartbeats [1] and is a valuable marker for cardiovascular control and respiratory function [2]. The chest strap may require shaving the chest hair and moistening the sensor to improve contact with the skin of the chest wall [3], while there is no specific preparation for the finger probe as long as the properly hydrated subject's finger has clean and dry skin. Hence, the importance of accurate and convenient measurement devices, the CorSense finger probe that is easier to use warrants a comparison to the Polar H7 chest strap considered the gold standard in our study [4]. This assessment occurred over three replicate measurements, during which HRV was recorded with both devices among 33 subjects aged 59-79 with coronary artery disease. The goal was to assess if finger probes could be a reliable substitute for conventional chest straps. Previous studies have corroborated both chest straps and finger probes for HRV monitoring. While the Polar H7 tool is consistent with ECG [4], evidence points to the photoplethysmography (PPG)-based finger probes as reliable alternatives [5][6]. Thus, our study seeks to contribute to this growing body of evidence.

2. Material and Methods

2.1. Raw Material

Our data set was collected in a clinical study (NCT04121741) [7]. The parent study aimed to collect data from up to 65 subjects with coronary artery disease between 59 and 79 years old who participated in three sessions. Subjects participated in two singing sessions and one non-singing control session. During the singing sessions, they sang along with music. HRV was measured simultaneously pre-, during-, and after-singing for 3 minutes each with the chest and finger probe during singing sessions. The non-singing control session recorded the same measurements but without singing. Subjects underwent a hearing test at the non-singing control visit. HRV was measured simultaneously pre-, during-, and after-hearing tests for 3 minutes each with the chest and finger probe during non-singing control sessions. The order of these three sessions was randomly assigned in a cross-over trial design. Patient demographic information and medical history were collected from the medical record. Recording

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.438

Table 1. HRV measurement.

HRV	Unit	Definition
rMSSD	ms	Root mean square of successive RR interval differences
SDNN	ms	Standard deviation of NN intervals
PNN5	%	Percentage of successive RR intervals that differ by more than 50 ms
LF Power	ms^2	Absolute power of the low-frequency band (0.04-0.15 HZ)
HF Power	ms^2	Absolute power of the high-frequency band (0.15-0.4 HZ)

initiation time did not differ by more than a couple of seconds between devices. Both devices transferred data to Elite HRV software [8]. This software allowed us to analyze the data and calculate more HRV variables. Finally, all data was stored in RedCAP. RedCAP enforced rules to reduce human error, such as min/max range checks [9].

2.2. Data Preparation

For this study, we exported the de-identified data from RedCAP. Based on the number of subjects who had HRV measured with both the chest strap and figure probe, we had data from 33 subjects. Each patient had nine sets of data, consisting of three visits and pre-, during, and after-singing during singing sessions and -hearing tests during control sessions. We kept only the non-singing control sessions to minimize measurement error since there were more discrepancies in singing sessions, possibly due to movement or sweat. Considering 33 subjects and three replicates for each (pre-, during, and after-hearing tests), we have 99 observations. For simplicity and precision, we restricted attention to the five most commonly used HRV measurements [10](Table 1). Originally, we had data from 33 subjects, or 99 observations, considering three replicates pre-, during- and after-hearing tests. The software from both devices informed us about poor-quality data. We removed 45% of the data for poor quality of the chest strap measurement and 9% for poor quality of the finger probe measurement. Just 1% deemed poor quality for both devices. Considering all the removed data, we have 46 observations. We also applied 10 percent trimming method, removing outliers to improve the results.

2.3. Linear regression

We analyzed HRV in both the time domain (rMSSD, SDNN, and PNN50) and the frequency domain (LF Power and HF Power). Linear regression is illustrated before and after trimming (Formula 1, Figure 1 and 2). Where $\boldsymbol{y} = [y_1, \ldots, y_n]$ is HRV data measured by finger probe CoreSense. And $\boldsymbol{x} = [x_1, \ldots, x_n]$ is HRV data measured by chest strap Polar H7. The purpose of linear regression is to see if there is a strong correlation between the two paired chest strap and finger probe measurements. HRV data from 33 subjects (mean age 69 with standard deviation 10, 52% female) were compared by data from both devices using linear regression. We made graphs to visualize linear regression of the data from the chest probe versus the finger probe for the data with and without outliers. Data shown in the plots were shown with the original scaling. The ideal results would be fitted exactly in a 45-degree line ($b = 1$). We calculated the linear regression error by mean squared error (MSE) (Formula 2). Where n is the number of data points, y_i is the actual value of the i_{th} data point, $\hat{y}_i$ is the predicted value of the i_{th} data point. The predicted value $\hat{y}_i$ is obtained using the model on the normalized data, with min-max scaling. For evaluating MSEs, we found normalized root-mean-square deviation (NRMSD) (Formula 3), where $\bar{y}$ is $(1/n)\sum_i y_i$. We also found the Coefficient of determination (R^2) for the paired chest strap and finger probe data (Formula 4).

$$\boldsymbol{y} = a + b\boldsymbol{x} \tag{1}$$

$$MSE = (1/n)\sum_i (y_i - \hat{y}_i)^2 \tag{2}$$

$$NRMSD = \sqrt{MSE}/\bar{y} \tag{3}$$

$$R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2 \tag{4}$$

2.4. Nonparametric methods

The Wilcoxon-signed-rank test (WSRT) is a nonparametric statistical method to compare two paired or dependent samples. Given appropriate conditions, this test could be applied in our study comparing HRV measurements between the chest strap and finger probe device. We had some ties (equal observations) in our data since the number of them was few (1-2); applying the Wilcoxon test and getting an approximation for them was acceptable. We also manually checked those ties in RedCAP to ensure they were not human errors. The WSRT was performed on the paired chest strap and finger probe data measurements of both devices to perform the comparison. We evaluated the WSRT results with p-value to determine whether the paired data are from the same distribution. We implemented the process of preparing data, linear regression method, and WSRT in Python to assess the agreement between HRV variables measured by chest strap and finger probe.

3. Results

Results for linear regression are shown in the plots separately for rMSSD, PNN50, SDNN, LF Power, and HF Power in their actual scaling (Figure 1 and 2). After trimming, the slope coefficients (b's) are found for linear regression lines to compare with the 45-degree line at each plot. MSE, NRMSD, and R^2 for rMSSD, PNN50, SDNN, LF Power, and HF Power were calculated to evaluate the linear regression method (Table 2). Figure 1 shows linear regression results and how trimming improved linear regression results for time-domain HRVs. Figure 2 shows linear regression results and how trimming improved linear regression results for frequency-domain HRVs. After trimming, NRMSD amounts were decreased by 12% for rMSSD, 67% for PNN50, 34% for SDNN, 81% for LF Power, and 45% for HF Power.

WSRT was repeated five times, considering five HRV measurements (rMSSD, SDNN, PNN50, LF Power, and HF Power). P-values were calculated to evaluate the WSRT method (Table 2).

Linear regression analyses and the WSRT showed closer correlations between devices for SDNN (0.212 NRMSD, 0.769 R^2, and 0.048 p-values) and LF Power (0.266 NRMSD, 0.958 R^2, and 0.047 p-values). However, none of them met the agreement criteria.

4. Discussion and Conclusions

Although R^2 values (related to linear regression) look good ($R^2 > 0.75$ for LF Power, PNN50, HF Power and SDNN), as a matter of NRMSD (related to linear regression) and p-values (related to WRST), the paired chest strap and finger probe data sets measured with the chest strap and the finger probe had significant differences. SDNN and LF Power had the closest p-values to 0.05. But they both were below the cutoff. In conclusion, none of the paired chest strap and finger probe HRV measurements met the agreement criteria even after 10% trimming. Although the finger probe is more convenient, the chest strap appears more accurate and has been validated with ECG. More significant discrepancies during the singing and the recovery period may be related to movement while singing, sweat affecting sensor contact with skin, or medical comorbidities such as peripheral vascular disease, which can be explored in future analyses.

Table 2. Linear and non-linear methods evaluation

Component [unit]	Linear regression R^2	Linear regression NRMSD	WSRT p-value
LF Power [ms^2]	0.958	0.266	0.047
PNN50 [u%]	0.794	0.349	0.010
HF Power [ms^2]	0.778	0.493	0.013
SDNN [ms]	0.769	0.212	0.048
rMSSD[ms]	0.662	0.353	0.008

References

[1] Malik M. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use: Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. Annals of Noninvasive Electrocardiology 1996;1(2):151–181.

[2] Lombardi F, Stein PK. Origin of Heart Rate Variability and Turbulence: An Appraisal of Autonomic Modulation of Cardiovascular Function. Frontiers in Physiology 2011; 2:95.

[3] Nigusse AB, Mengistie DA, Malengier B, Tseghai GB, Langenhove LV. Wearable Smart Textiles for Long-

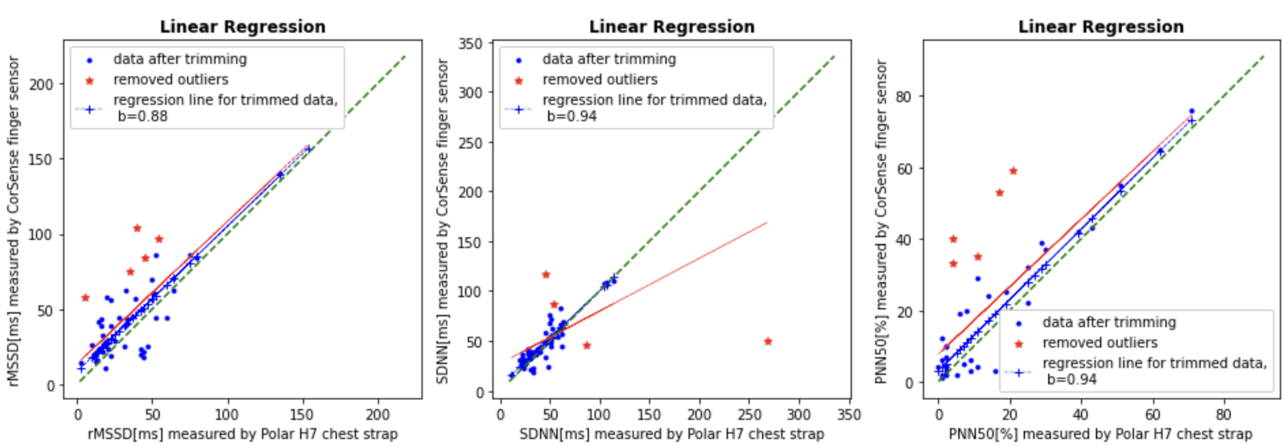

Figure 1. Linear regression for time domain HRV data (rMSSD, SDNN, and PNN50).

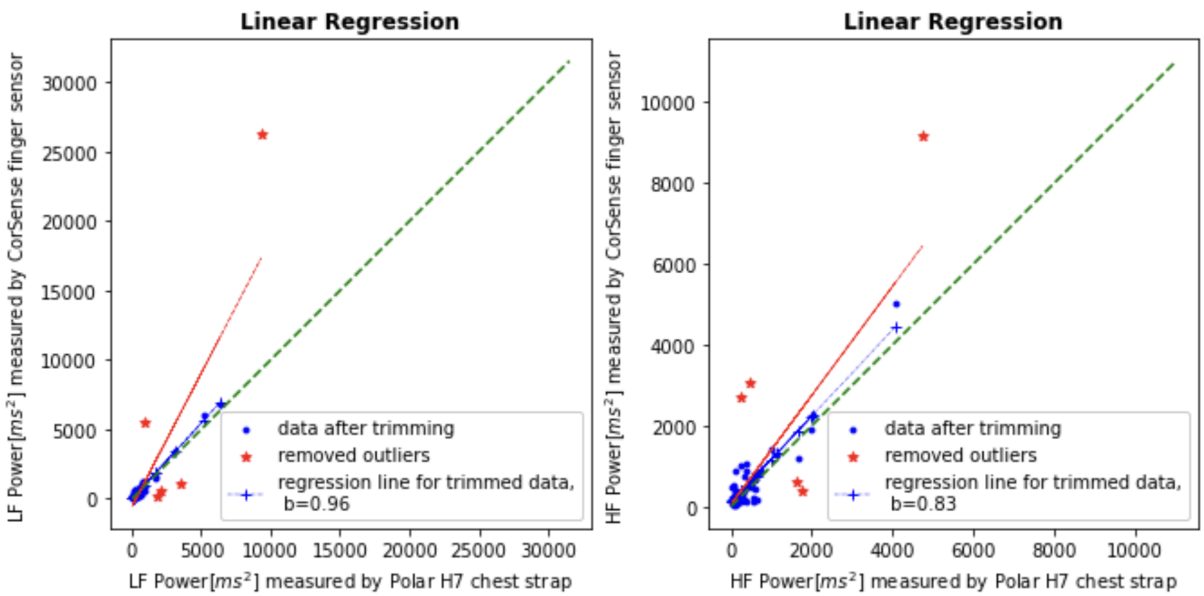

Figure 2. Linear regression for frequency domain HRV data (LF Power, and HF Power).

term Electrocardiography Monitoring—A Review. Sensors 2021;21(12):4174.

[4] Daniel J Plews Ben Scott MAMWAEKPBL. Comparison of Heart-Rate-Variability Recording with Smartphone Photoplethysmography, Polar H7 Chest Strap, and Electrocardiography. International Journal of Sports Physiology and Performance Nov. 2017;12(10):1324–1328.

[5] Kinnunen H, Rantanen A, Kenttä T, Koskimäki H. Feasible Assessment of Recovery and Cardiovascular Health: Accuracy of Nocturnal HR and HRV Assessed via Ring PPG in Comparison to Medical Grade ECG. Physiological Measurement 2020;41(4):04NT01.

[6] Speer KE, Semple S, Naumovski N, McKune AJ. Measuring Heart Rate Variability Using Commercially Available Devices in Healthy Children: A Validity and Reliability Study. European Journal of Investigation in Health Psychology and Education 2020;10(1):390–404.

[7] Kulinski J. Evaluating the Impact of Singing Interventions on Markers of Cardiovascular Health in Older Patients with Cardiovascular Disease (NCT04121741; Version V1) [Data Set] 2020;URL `https://clinicaltrials.gov/ct2/show/NCT04121741`.

[8] Elite HRV. URL `https://elitehrv.com/`.

[9] Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, et al. The REDCap Consortium: Building an International Community of Software Platform Partners. Journal of Biomedical Informatics 2019;95:103208.

[10] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health 2017; 258.

Address for correspondence:

Mehri Bagheri-Mohamadi-Pour
University of Wisconsin-Milwaukee,
Milwaukee, WI, USA
Mehri@uwm.edu