

Original Paper

# Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study

Emily Lam<sup>1\*</sup>, BEng; Shahrose Aratia<sup>2\*</sup>, BSc; Julian Wang<sup>3</sup>, MD; James Tung<sup>4</sup>, PhD

<sup>1</sup>Possibility Engineering and Research Laboratory, Bloorview Research Institute, Toronto, ON, Canada

<sup>2</sup>Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Michael G Degroote School of Medicine, McMaster University, Hamilton, ON, Canada

<sup>4</sup>Department of Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada

\*these authors contributed equally

**Corresponding Author:**

James Tung, PhD

Department of Mechanical & Mechatronics Engineering

University of Waterloo

200 University Ave W

Waterloo, ON, N2L 3G1

Canada

Phone: 1 519 888 4567 ext 43445

Email: [james.tung@uwaterloo.ca](mailto:james.tung@uwaterloo.ca)

## Abstract

**Background:** Heart rate variability (HRV) is used to assess cardiac health and autonomic nervous system capabilities. With the growing popularity of commercially available wearable technologies, the opportunity to unobtrusively measure HRV via photoplethysmography (PPG) is an attractive alternative to electrocardiogram (ECG), which serves as the gold standard. PPG measures blood flow within the vasculature using color intensity. However, PPG does not directly measure HRV; it measures pulse rate variability (PRV). Previous studies comparing consumer-grade PRV with HRV have demonstrated mixed results in short durations of activity under controlled conditions. Further research is required to determine the efficacy of PRV to estimate HRV under free-living conditions.

**Objective:** This study aims to compare PRV estimates obtained from a consumer-grade PPG sensor with HRV measurements from a portable ECG during unsupervised free-living conditions, including sleep, and examine factors influencing estimation, including measurement conditions and simple editing methods to limit motion artifacts.

**Methods:** A total of 10 healthy adults were recruited. Data from a Microsoft Band 2 and a Shimmer3 ECG unit were recorded simultaneously using a smartphone. Participants wore the devices for >90 min during typical day-to-day activities and while sleeping. After filtering, ECG data were processed using a combination of discrete wavelet transforms and peak-finding methods to identify R-R intervals. P-P intervals were edited for deletion using methods based on outlier detection and by removing sections affected by motion artifacts. Common HRV metrics were compared, including mean N-N, SD of N-N intervals, percentage of subsequent differences >50 ms (pNN50), root mean square of successive differences, low-frequency power (LF), and high-frequency power. Validity was assessed using root mean square error (RMSE) and Pearson correlation coefficient ( $R^2$ ).

**Results:** Data sets for 10 days and 9 corresponding nights were acquired. The mean RMSE was 182 ms (SD 48) during the day and 158 ms (SD 67) at night.  $R^2$  ranged from 0.00 to 0.66, with 2 of 19 (2 nights) trials considered moderate, 7 of 19 (2 days, 5 nights) fair, and 10 of 19 (8 days, 2 nights) poor. Deleting sections thought to be affected by motion artifacts had a minimal impact on the accuracy of PRV measures. Significant HRV and PRV differences were found for LF during the day and R-R, SDNN, pNN50, and LF at night. For 8 of the 9 matched day and night data sets,  $R^2$  values were higher at night ( $P=.08$ ). P-P intervals were less sensitive to rapid R-R interval changes.

**Conclusions:** Owing to overall poor concurrent validity and inconsistency among participant data, PRV was found to be a poor surrogate for HRV under free-living conditions. These findings suggest that free-living HRV measurements would benefit from examining alternate sensing methods, such as multiwavelength PPG and wearable ECG.

(JMIR Biomed Eng 2020;5(1):e17355) doi: [10.2196/17355](https://doi.org/10.2196/17355)

**KEYWORDS**

heart rate determination photoplethysmography; wearable electronic sensors; physiological monitoring; ambulatory monitoring; mobile phone

## Introduction

### Motivation

With the growing ubiquity of commercially available wearable technologies, obtaining long-term physiological measurements under free-living conditions is feasible and permits longitudinal examination of ecologically valid patterns. This presents an opportunity for continuous patient monitoring under free-living conditions, including the potential to identify at-risk individuals (eg, patients with cardiac disease). Heart rate variability (HRV) is a well-established, powerful metric used to assess cardiac health, including autonomic nervous system function regulating cardiac activity. Compared with an individual's heart rate (HR) averaged over a short period, HRV measures variations in HR primarily as an indicator of the efforts of the sympathetic and parasympathetic nervous systems to achieve an optimal cardiac response under constantly changing stimuli [1]. Previous research has explored the use of HRV monitoring in predicting or detecting sleep quality [2], mental stress [3], chronic pain [4], posttraumatic stress disorder [5], bipolar disorder [5], and cardiac health [6].

### Measuring HRV

The (gold) criterion standard for measuring HRV is through an electrocardiogram (ECG) to obtain a direct recording of cardiac electrical activity. On ECG, the R wave represents the maximum upward deflection of a normal QRS complex. The duration between two successive R waves defines the R-R interval [7], which is used to measure HR and HRV. Although wearable ECGs exist, they typically require electrodes affixed to the skin, which makes them obtrusive and can cause skin breakdown, and they are also prone to motion artifacts during day-to-day activities [8]. Alternatively, photoplethysmography (PPG) uses an optical sensor widely used to unobtrusively track mean HR, especially in wrist-worn devices (eg, Fitbit).

### PPG for Pulse Rate Variability

PPG sensors measure changes in pulsatile blood flow within an individual's vasculature using color intensity signals [9]. Signal peaks associated with the flow of blood are used as indicators of HR, allowing for the calculation of peak-to-peak (P-P) intervals. PPG sensors do not directly measure HRV; instead, they measure pulse rate variability (PRV), the change in vessel pulse periods, from which P-P intervals denote a pulse rate (PR) [10]. PPG sensors can be placed at a variety of measurement sites including the fingers, wrist, brachia, ear, forehead, and esophagus without requiring additional equipment. This makes PPG especially convenient for pervasive cardiac monitoring [11], with well-validated use for mean HR measurements [4]. Although evidence examining PPG capabilities to accurately measure HRV shows promise, studies comparing PPG with gold standard ECG methods under free-living conditions remain limited.

The accuracy of PRV as a measure of HRV has been investigated with clinical devices under controlled, and often stationary, conditions [12-17]. Although these studies indicate that PRV may be a useful as a proxy measure of HRV using medical-grade devices under controlled conditions, studies using wearable consumer-facing devices have shown mixed results. These few studies largely use short-term collections in controlled circumstances, some of which do not simultaneously collect ECG [18,19]. A systematic review by Georgiou et al [20] found that wearable devices can provide accurate measurement of HRV measures at rest; however, accuracy declines as exercise and motion levels increase. The review also showed that heterogeneity in sensor position, detection algorithm, experimental settings, and analysis methods from existing studies limits the evidence. A review by Shäefer and Vagedes [21] found similar results, suggesting that physical activity and mental stressors lead to unacceptable deviations between PRV and HRV. Ultimately, further research is required to determine the efficacy of PRV in estimating HRV during free-living conditions in which individuals are unrestricted and engaging in their daily activities [22].

### Limitations of PPG

PPG sensors have been found to be sensitive to motion artifacts, changes in blood flow caused by movement, compression and deformation of the vasculature arising from pressure disturbances at the interface between the sensor and the skin [11], and light leaking between the sensor and the skin [23]. Some studies have examined the removal of motion artifacts from PPG signals using signal processing techniques and acceleration as a reference [23-28]. For example, methods involving accelerometry have shown promise for improving coherence by editing signals likely influenced by motion artifacts [28,29]. Baek and Shin [30] collected PPG measurements over 24 hours using a custom device and filtering method, recommending a subset of HRV metrics as good targets for continuous HRV tracking using commercial devices. Morelli et al [28] conducted a study evaluating the accuracy of a consumer-grade PPG (Microsoft Band 2) for HRV estimation during less restrictive, but controlled, conditions (eg, sitting and walking) over 10-min trials. Errors likely caused by motion artifacts during walking were attenuated by using corresponding accelerometer signals to delete sections of the data corrupted by motion artifacts.

### Objectives

Although HR and PR are correlated and closely related, the use of PRV to estimate HRV requires further research, especially under free-living conditions. In this study, the concurrent validity of PRV measurements from a consumer-facing PPG sensor is compared with HRV measurements from a portable ECG under 2 unsupervised conditions up to 4.5 hours each: (1) while engaging in regular activities of daily living and (2) during sleep. A secondary goal of this study is to examine factors

influencing estimation errors of PRV for HRV, including motion artifacts, measurement conditions, and editing approaches.

## Methods

### Participants

A convenience sample of healthy individuals aged 18-65 years was recruited for the study. Individuals with a history of cardiac and/or sleep disorders were excluded to minimize the collection of irregular cardiac signals. Under these conditions, approval for this study was granted by the University of Waterloo Research Ethics Committee on September 5, 2017, filed under protocol #31197.

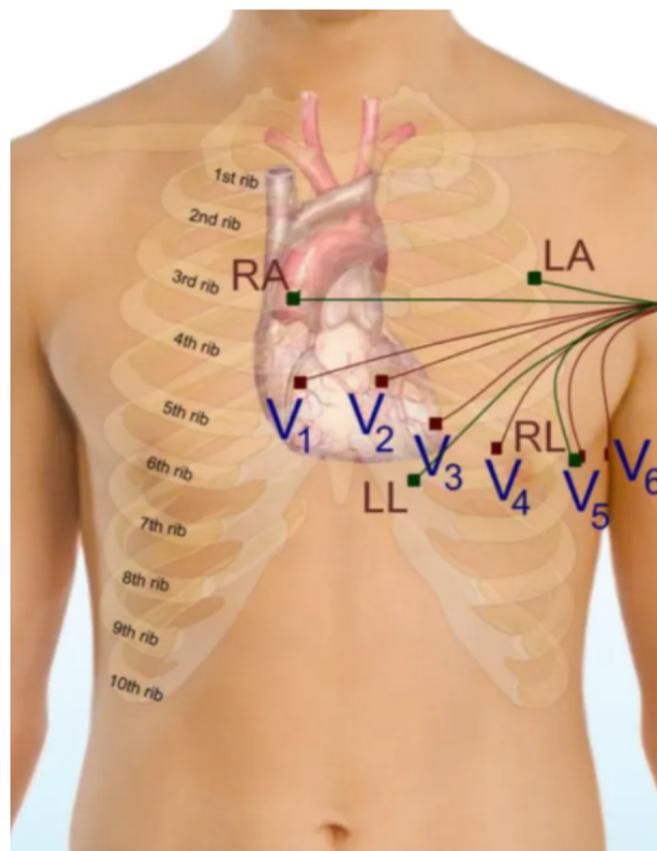
### Device Setup

A total of 2 wearable devices were used to acquire cardiovascular signals in this study: (1) a commercially available optical PPG wearable device (Microsoft Band 2 or MB2, Microsoft) and (2) a research-grade wearable ECG device (Shimmer3 ECG, Shimmer). Both wearables were recorded simultaneously with signals transmitted via Bluetooth to a smartphone (Pixel or Nexus 3, Google). To synchronize the

devices, triaxial accelerations were also recorded with both devices. Participants were asked to wear the devices twice, for at least 90 min each, once during daily activities and a second time when sleeping.

To record ECG, hydrogel electrodes (Kendall 233 Hydrogel, Covidien) were placed in a 4-lead bipolar limb lead configuration (ie, left arm [LA], right arm [RA], left leg [LL], right leg) on the participant's chest as shown in Figure 1 [31]. The participant's skin was prepared by shaving and sanitized using hospital-grade alcohol wipes before electrode placement. Electrodes were connected to a Shimmer3 ECG, worn at the waist with a strap, and all leads were taped to the chest to prevent tangling and static, and minimize motion artifacts. On the smartphone, the Multi-Shimmer Sync mobile app (Shimmer, Dublin, Ireland) was used to record ECG data from the Shimmer3 ECG. The MB2 was worn on the participant's wrist of choice as tightly as possible, without causing discomfort. MB2 size (small or medium) was selected to fit the size of the participant's wrist. A third-party mobile app, Companion for Microsoft Band (released by Pain in My Processor, Google Play Store), was used to log data from the MB2 to the smartphone.

**Figure 1.** Electrocardiogram 4-lead bipolar limb electrocardiogram configuration on participants' chests. LA: left arm; RA: right arm; V: precordial leads.



### Participants' Instructions

Given the free-living nature of data collection, participants were instructed on how to set up and monitor device connection and logging status to facilitate troubleshooting. To ensure proper electrode placement, a (trained) researcher placed the electrodes in the 4-lead bipolar limb lead configuration (Figure 1) during

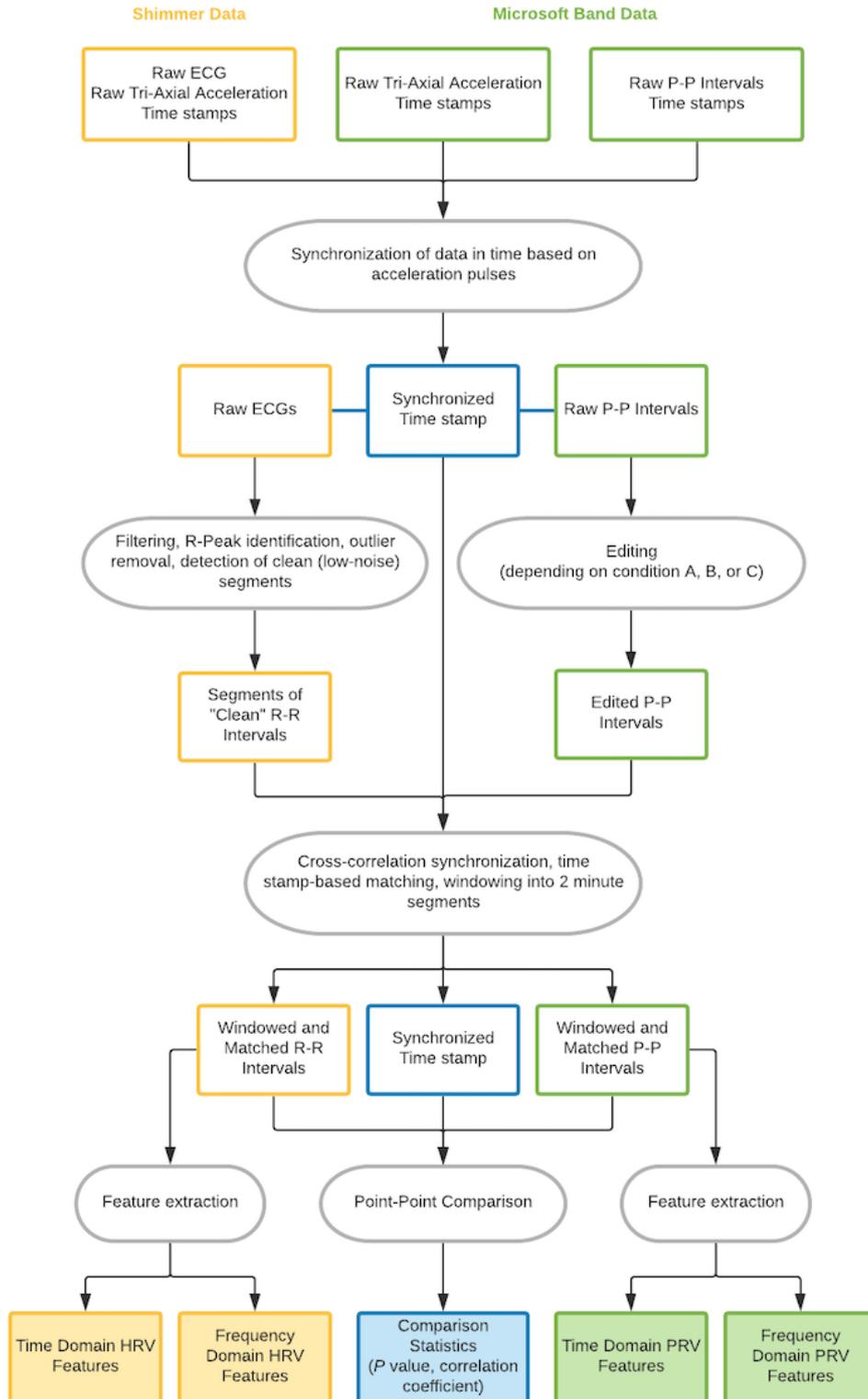
the first data collection (ie, day). For the second collection (ie, night), electrodes were left on, replaced by the research assistant, and/or marked by location and replaced by the participant. Before the second collection, ECG signals were visually examined to ensure that the QRS complexes were clearly identifiable. Participants were instructed and encouraged to

contact a researcher at any time in case of questions or concerns during data collection.

**Postprocessing**

Following data collection, all postprocessing and statistical analyses were conducted using MATLAB 2018a (MathWorks). Figure 2 outlines the steps taken in postprocessing.

**Figure 2.** Postprocessing of data from the Shimmer and Microsoft Band. ECG: electrocardiogram; HRV: heart rate variability; P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; PRV: pulse rate variability; R-R: time between 2 R peaks in an ECG.



**Synchronizing Devices**

Shimmer3 and MB2 were coarsely synchronized by aligning triaxial acceleration peaks from tapping both devices

simultaneously on a table. Each device was tapped 3 times in 2 orientations with 10 s of rest between orientations. Fine

synchronization was performed using a cross-correlation method described below (cross-correlation synchronization).

### ECG Data Processing

Both LA-RA and LL-RA ECG signals were filtered using a first order bandpass Butterworth filter from 1 to 25 Hz. A maximal overlap discrete wavelet transform with a Daubechies least-asymmetric wavelet with 4 vanishing moments was used to enhance the R peaks in the ECG, followed by a threshold-based peak-finding function used to identify the R-peaks [32,33]. In one sample (participant 2, daytime), the wavelet detection algorithm more accurately and consistently detected the T wave of the ECG signal and was used as a proxy for the QRS complex, previously shown to give results similar to those of R-peak detection [34]. A time series of R-R intervals was extracted from the detected R-peaks, and outlier values outside of the physiological range of values for a healthy individual at rest, walking, or during sleep ( $R-R < 0.3$  s or  $R-R > 2.5$  s) were removed [35-37]. To remove transients associated with artifacts or noise, segments of at least 15 consecutive R-R intervals were included in the analysis. Longer segment thresholds (30, 60, and 120 consecutive R-R intervals) were tested with negligible effects on the results. On the basis of the signal that provided more R-R intervals, either the LA-RA or LL-RA electrode pair signal was chosen for processing and analysis.

### PPG Data Processing

P-P intervals and corresponding time stamps were recorded directly from MB2 outputs as the time interval between 2 continuous heartbeats [38]. Note that the temporal resolution of MB2 is limited to 10 ms. On the basis of existing literature reporting signal processing methods to edit R-R intervals and remove artifacts, three methods were used to identify and delete artifacts in the P-P intervals outputted by the MB2, resulting in 4 conditions of P-P data. Deletion was chosen as the editing technique (as opposed to interpolation) because motion artifacts would likely affect consecutive samples, making interpolation challenging. In addition, the long-term nature of data collection would mitigate one of the major concerns associated with deletion, the loss of samples [39]. The 4 processing conditions were as follows:

- None (condition A): This condition contains the raw P-P intervals.
- Threshold deletion (condition B): Removing implausible P-P interval values for a healthy individual at rest, walking, or sleeping ( $P-P < 0.3$  s or  $P-P > 2.5$  s) [35-37].
- Moving average deletion (condition C): Threshold deletion (as described in B above) and removing changes in P-P intervals faster than physiologically plausible indicated by a moving average filter. This was done following Morelli et al [28], discarding values for which  $|PP_t - \mu_{10}| \geq 0.5\mu_{10}$ , where  $PP_t$  refers to the P-P interval data and  $\mu_{10}$  is a 10 s moving average.
- Acceleration-based deletion (condition D): A series of threshold filters, moving average filters (described in C above), and an acceleration filter. Considering that low PPG signal quality may be attributable to movement, Morelli et al [28] removed signal segments affected by motion artifacts

by estimating periods of signal quality associated with the corresponding accelerometry time series,  $W_t$ , and then removing P-P intervals where  $W_t$  was found to exceed a threshold,  $k$ .  $k$  was identified by examining the correlation between  $W_t$  and error, where  $W_t$  is calculated as an average of  $w_t$  over a window of duration  $\tau$ , and  $W_t$  is calculated as follows:

$$W_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2 + (z_t - z_{t-1})^2}$$

In this study, no significant correlation between  $W_t$  and  $\tau$  was found. As such, a threshold of  $\tau = 0.02$  m/s<sup>2</sup> was used to filter the data with  $\tau = 40$  s (the same parameters as used by Morelli et al [28]).

### Data Synchronization

Following coarse synchronization of MB2 and Shimmer3, consistent delays between the 2 devices were observed. To identify the highest correlation between devices, a cross-correlation between P-P and R-R data was conducted. The estimate of the time-shift was applied to the P-P data, similar to the method used by Pietilä et al [40]. P-P intervals were then matched to R-R intervals by matching data points with the closest time stamps. If a data point did not have a matching interval within 1 s, the interval was deleted. The 1-s delay was chosen to accommodate for delays in Bluetooth transmission and pulse transit time. After matching, the remaining data were divided into 2-min windows from which the HRV and concurrent validity metrics were calculated [41].

### HRV Metrics for Analysis

After postprocessing, the following time domain HRV and PRV features were extracted for each trial, where N-N refers to either R-R or P-P:

- Mean N-N: the mean of all N-N intervals
- Mean HR: reciprocal of mean N-N, in beats per minute (bpm)
- SDNN: a measure of overall variability, the SD of all N-N intervals, also known as RRSD
- pNN50: percentage of subsequent differences more than 50 ms
- RMSSD: root mean square of subsequent differences
- LF, HF, LF/HF ratio: low-frequency power (LF), high-frequency power (HF), and the ratio of LF to HF
- SD1 and SD2: SDs of short ( $x=y$ ) and long (orthogonal to  $x=y$ ) diagonal Poincaré plot axes [12]

For spectral measures, R-R and P-P intervals were converted to instantaneous HR ( $60/N-N$ , where N-N is interval time in seconds) and then interpolated to 4 Hz using a piecewise cubic Hermite interpolation (MATLAB function "pchip"). This ensured regular time intervals between data points, a prerequisite for estimating the Fourier transform and signal power. The Fourier transform was performed (using "fft" function in MATLAB) on the entire data set for each participant. This allowed for the calculation of frequency domain HRV features such as LF (0.04-0.15 Hz) and HF (0.15-0.40 Hz). LF and HF

were computed in normalized units by the sum of LF and HF. The ratio of LF to HF was also reported.

### Analyses

To quantify the concurrent validity between R-R and P-P intervals, the following metrics were used:

- Root mean square error (RMSE): RMSE between matched R-R and P-P samples
- Pearson correlation coefficient ( $R^2$ ): The correlation strength between R-R and P-P intervals.  $R^2$  values were categorized as strong ( $R^2 \geq 0.7$ ), moderate ( $0.5 \leq R^2 < 0.7$ ), fair ( $0.3 \leq R^2 < 0.5$ ), and poor ( $0.3 < R^2$ ) [42].

To compare PPG-derived metrics across collection and processing conditions (ie, day- or nighttime collection, filtering condition), two-tailed paired  $t$  tests were used. Bland-Altman plots were generated to illustrate the agreement between R-R and P-P intervals. In the Bland-Altman plot, the difference between each P-P and R-R measurement is plotted against the mean of each measurement [43].

## Results

### Overview

This section presents the results of (1) investigating the concurrent validity between R-R and P-P intervals across published filtering methods, (2) a comparison between ECG- and PPG-derived metrics of HRV, and (3) a comparison across free-living data collection conditions (ie, day and night). A total of 10 volunteers were recruited (3 men and 7 women, aged 20-61 years) for this study for a total of 19 trials (1 day and 1 night per participant). One participant's ECG night data were corrupted and therefore not analyzed or reported.

After processing, a large amount of data was lost. The number of matched and windowed N-N intervals is described in [Table 1](#); all comparison statistics were calculated on the basis of these data. The percentages of compared intervals were calculated by dividing the number of matched and windowed samples by the total number of R-R or P-P intervals detected from the ECG or MB2, respectively.

**Table 1.** Group mean (SD) of data sample sizes used for comparison between R-R and P-P intervals across processing and collection conditions.

Collection condition and processing condition	Number of samples, mean (SD)	Percent R-R intervals compared, mean (SD)	Percent P-P intervals compared, mean (SD)
<b>Day</b>			
A	5168.70 (1683.92)	52.35 (26.89)	47.29 (18.32)
B	4706.5 (1447.57)	48.25 (26.30)	43.91 (18.31)
C	3311.30 (1316.13)	34.68 (24.73)	32.29 (16.88)
D	1847.40 (1334.28)	23.03 (26.28)	21.39 (15.92)
<b>Night</b>			
A	8418.78 (5179.41)	55.05 (27.70)	53.30 (19.26)
B	8197.11 (5060.21)	53.79 (27.31)	52.06 (19.15)
C	7383.00 (4075.76)	46.89 (24.24)	46.66 (19.38)
D	7177.33 (4901.63)	41.15 (27.67)	42.97 (21.23)

A larger data sample was acquired at night than that acquired during the day. Despite formal instructions and training on the operation and charging of the sensor systems, several technical barriers were frequently encountered that limited the number of samples in each trial. These included inadvertent misplacement of ECG electrodes or MB2, insufficient battery charging before night collection, and/or dropped Bluetooth stream to the mobile device.

### Concurrent Validity Across the Editing Techniques

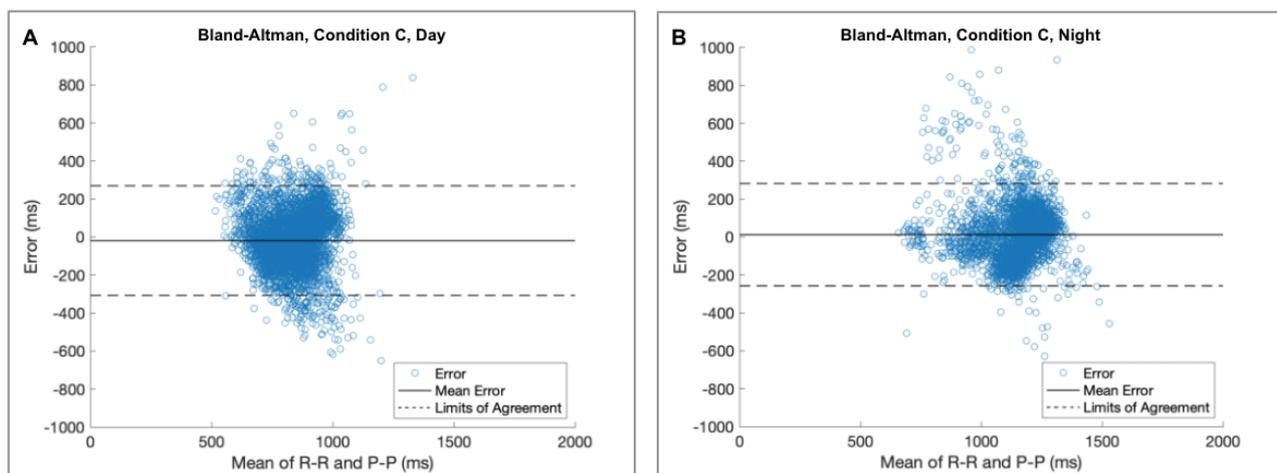
[Table 2](#) compares the concurrent validity of P-P data with that of the R-R data across all processing conditions, including RMSE and  $R^2$ . The largest differences were observed in the RMSE between the raw (A) and filtered (B, C, and D) conditions.

**Table 2.** Group mean root mean square error, concurrent validity ( $R^2$ ), and number of matched samples across processing and collection conditions.

Processing condition	Day (n=10)	Night (n=9)
<b>Root mean square error (ms), mean (SD)</b>		
A	182 (48)	158 (67)
B	165 (42)	136 (53)
C	144 (39)	120 (45)
D	122 (47)	119 (45)
<b><math>R^2</math>, mean (SD)</b>		
A	0.15 (0.12)	0.28 (0.17)
B	0.14 (0.13)	0.33 (0.19)
C	0.18 (0.13)	0.34 (0.21)
D	0.22 (0.17)	0.34 (0.21)

The RMSE ranged between 46 and 285 ms across all conditions. Increased editing reduced the average error (RMSE). Under condition C, error was further examined by generating

Bland-Altman plots comparing the P-P intervals with R-R intervals, as shown in Figure 3. Although the mean error is close to zero for both day and night conditions, the limits of agreement were greater than 200 ms.

**Figure 3.** Bland-Altman plots for 1 participant under processing condition C for (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R: time between 2 R peaks in an electrocardiogram.

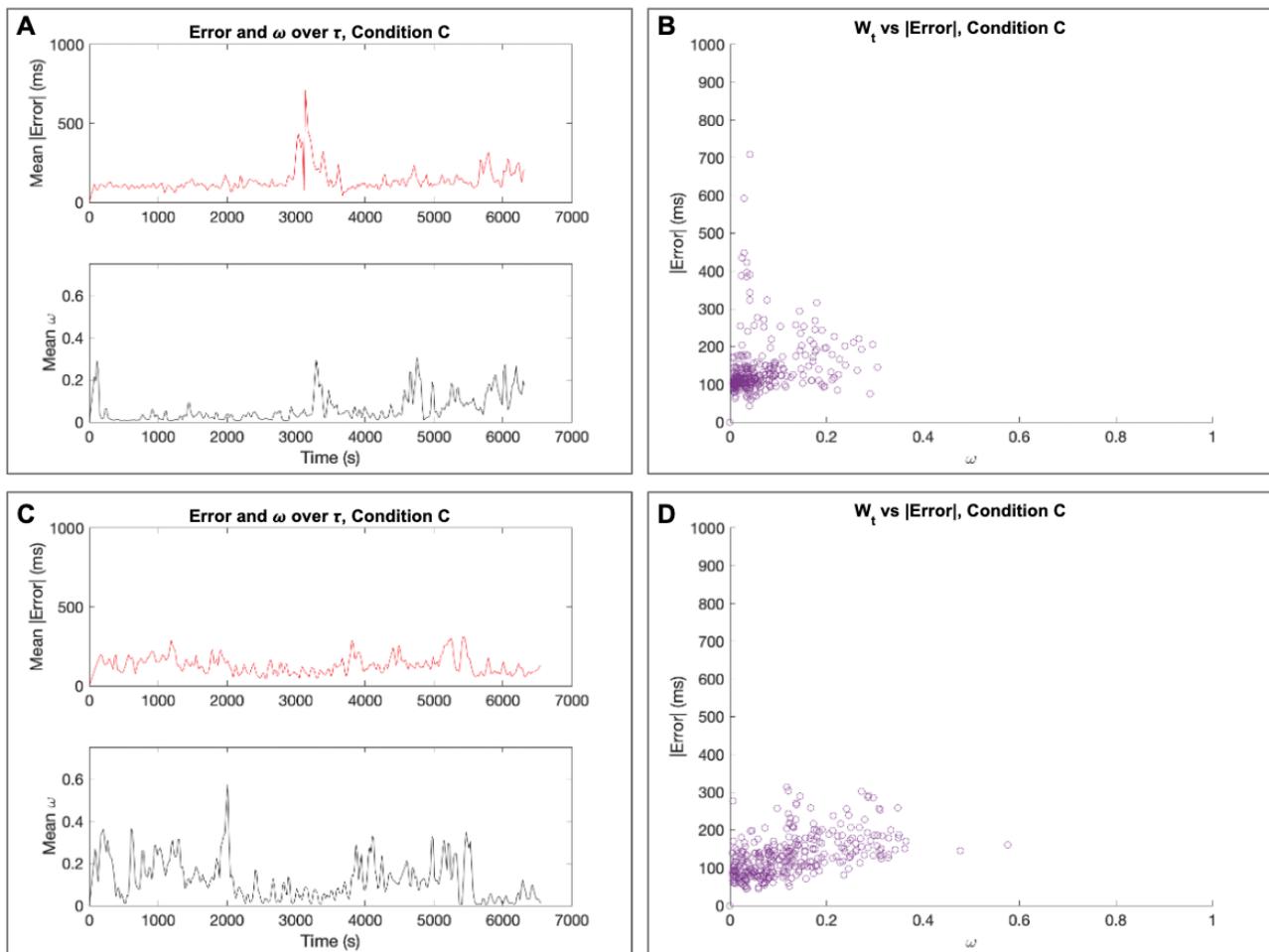
Across all conditions,  $R^2$  values ranged from 0 to 0.66. Editing did not have a large impact on  $R^2$ . Although  $R^2$  improved at night, none of the correlations were considered strong; 2 of 19 (all night) were moderate, 7 (2 days, 5 nights) were fair, and 10 (8 days, 2 nights) were poor. Of the 19, 16 (9 days, 7 nights) paired  $t$  tests between R-R and P-P intervals under condition C yielded  $P=.01$ , indicating significant differences between ECG- and PPG-based methods.

Under condition D, no data sets showed strong correlations. Only 3 (1 day, 2 nights) were moderate, 7 were fair (1 day, 6 nights), and 9 were poor (7 days, 2 nights). Paired  $t$  tests between matched R-R and P-P intervals edited under condition D were significant for 12 trials (5 days, 7 nights). Notably, condition

D reduced the amount of data available for analysis, especially during the day. From condition C to D, the average sample loss was 40.18% (SD 29.59) during the day and 3.73% (SD 4.37) at night.

Compared with condition C, condition D improved RMSE and  $R^2$  slightly during the day and varied by trial. The mean correlation between error and  $W_t$  was 0.28 (SD 0.24), with a range of 0.13 to 0.70 for day data, and 0.29 (0.21), with a range of 0.16 to 0.73 for night data. Figure 4 [44] shows the error and  $W_t$  for a sample showing lower correlation between  $W_t$  and error ( $R^2=0.16$ ) and a sample trial with higher observed correlation ( $R^2=0.50$ ).

**Figure 4.** Correlation between absolute error and mean change in triaxial acceleration ( $W_t$ ) under the same conditions as Morelli et al (A) and (B) comparison of  $|\text{Error}|$  and  $W_t$  over time for a sample with low correlation ( $R^2=0.16$ ) (C), and (D) comparison of  $|\text{Error}|$  and  $W_t$  over time for a sample with higher correlation ( $R^2=0.50$ ).



### Comparison of HRV and PRV Measures

Table 3 compares the HRV and PRV measures across participants under condition C, as this condition yielded the highest concurrent validity for most participants while retaining sample size. The findings in Table 3 are based on a 3311.30 (SD 1316.13) matched samples for day data and 7303.00 (SD 4075.76) for night data. Under condition C, paired  $t$  tests revealed no significant differences between HRV and PRV

measures. At night, SDNN, pNN50, RMSSD, SD1, SD2, LF, HF, and LF/HF ratio metrics were observed to be significantly different. Significant differences between HRV and PRV measures were observed in more measures at night, a condition during which motion artifacts are expected to be lower, allowing for collection of more accurate PRV data. Note that the temporal resolution of MB2 is limited to 10 ms, but many of the observed differences between R-R and P-P intervals are larger.

**Table 3.** Comparison of mean heart rate variability and pulse rate variability metrics under processing condition C.

Features	Day				Night			
	HRV <sup>a</sup>	PRV <sup>b</sup>	Error	<i>P</i> value <sup>c</sup>	HRV	PRV	Error	<i>P</i> value <sup>c</sup>
<b>Time domain features</b>								
NN (ms), mean (SD)	829 (70)	833 (51)	19 (15)	.59	967 (151)	960 (142)	10 (9)	.08
SDNN <sup>d</sup> (ms), mean (SD)	90 (36)	98 (25)	25 (20)	.48	87 (37)	69 (25)	20 (10)	.03
pNN50 <sup>e</sup> (%), mean (SD)	30.60 (24.51)	39.74 (16.18)	15.90 (11.30)	.14	38.58 (30.59)	21.35 (15.06)	19.48 (14.11)	.02
RMSSD <sup>f</sup> (ms), mean (SD)	104 (58)	116 (38)	42 (36)	.54	101 (57)	67 (20)	34 (36)	.02
SD1 <sup>g</sup> (ms), mean (SD)	74 (41)	82 (27)	30 (26)	.54	72 (40)	48 (21)	24 (26)	.02
SD2 <sup>h</sup> (ms), mean (SD)	94 (40)	110 (25)	31 (33)	.24	97 (35)	83 (29)	18 (14)	.05
<b>Frequency domain features</b>								
LF <sup>i</sup> (nu), mean (SD)	0.70 (0.03)	0.69 (0.02)	0.03 (0.02)	.43	0.70 (0.03)	0.72 (0.01)	0.02 (0.02)	.02
HF <sup>j</sup> (nu), mean (SD)	0.30 (0.03)	0.31 (0.02)	0.03 (0.02)	.43	0.30 (0.03)	0.28 (0.01)	0.02 (0.02)	.02
LF/HF ratio, mean (SD)	2.39 (0.32)	2.26 (0.22)	0.28 (0.23)	.29	2.43 (0.28)	2.64 (0.13)	0.21 (0.18)	.01

<sup>a</sup>HRV: heart rate variability.

<sup>b</sup>PRV: pulse rate variability.

<sup>c</sup>Results from paired *t* test between HRV and PRV measures.

<sup>d</sup>SDNN: SD of all N-N intervals.

<sup>e</sup>pNN50: percent of subsequent differences more than 50 ms.

<sup>f</sup>RMSSD: root mean square of subsequent differences.

<sup>g</sup>SD1: SD of short ( $x=y$ ) Poincaré plot axis.

<sup>h</sup>SD2: SD of long (orthogonal to  $x=y$ ) Poincaré plot axis.

<sup>i</sup>LF: low-frequency power.

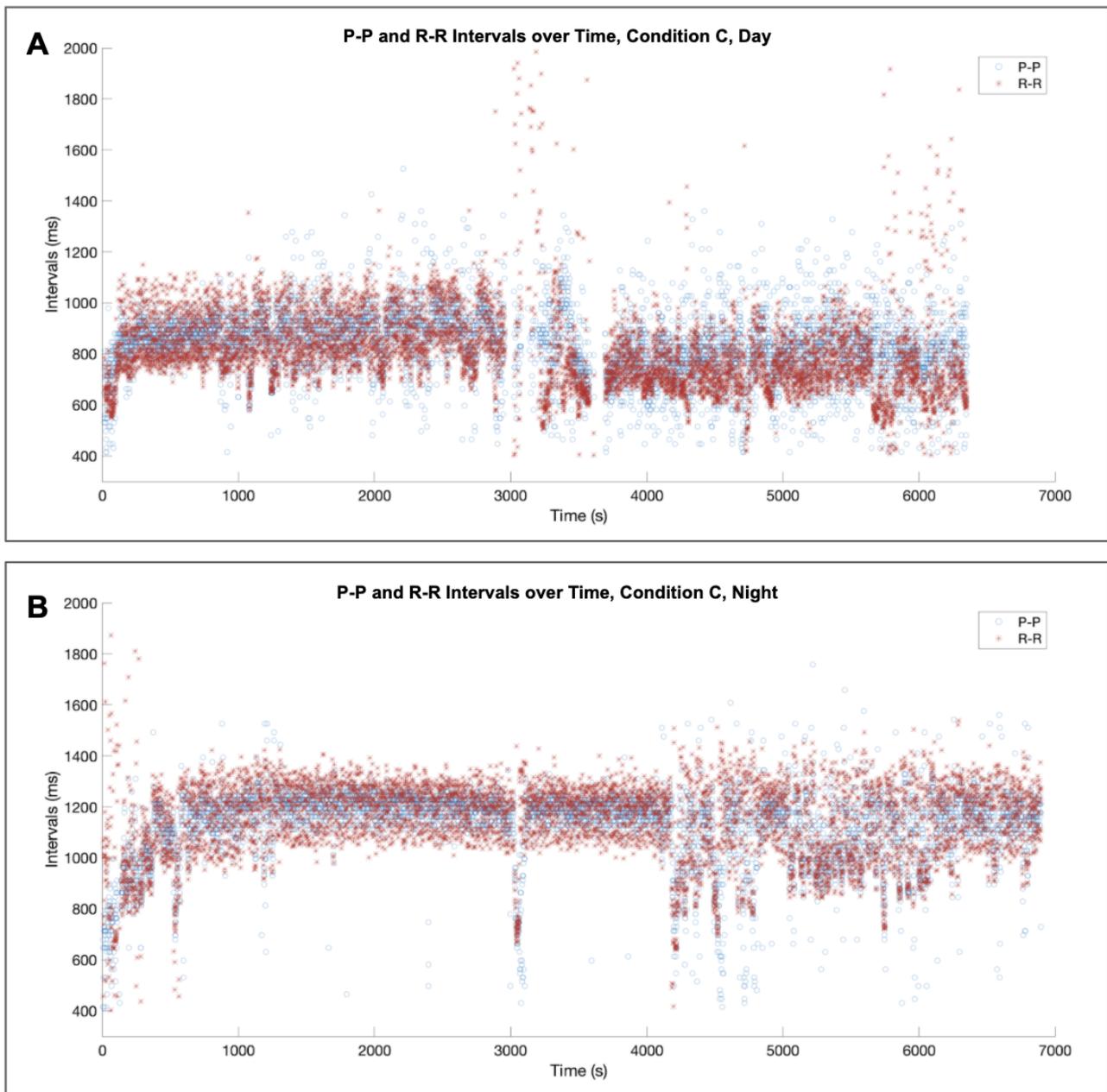
<sup>j</sup>HF: high-frequency power.

Compared with processing condition C, similar results were observed in condition D (Multimedia Appendix 1). Under condition D, paired *t* tests revealed significant differences between HRV and PRV measures for no measures during the day, but there were significant differences in R-R and pNN50 at night. Although this may be attributed to condition D using motion artifact editing, the large number of samples edited from condition C to D may partially explain these findings. Given the large sample loss associated with condition D and a lack of strong correlation between  $W_t$  and error, the remainder of this

study focuses on the results from processing condition C (over D).

Time series plots of matched and edited R-R and P-P intervals (Figure 5) highlight several differences between the ECG and PPG methods. Similar to the mean N-N results, the data sets follow the same trends on average, but there are notable differences. First, P-P intervals seem to be less sensitive to changes in R-R intervals, as many shorter and longer intervals were not well matched. Fewer artifacts were observed in the R-R intervals that did not appear in the P-P interval signal, which may be attributable to less R-R interval editing.

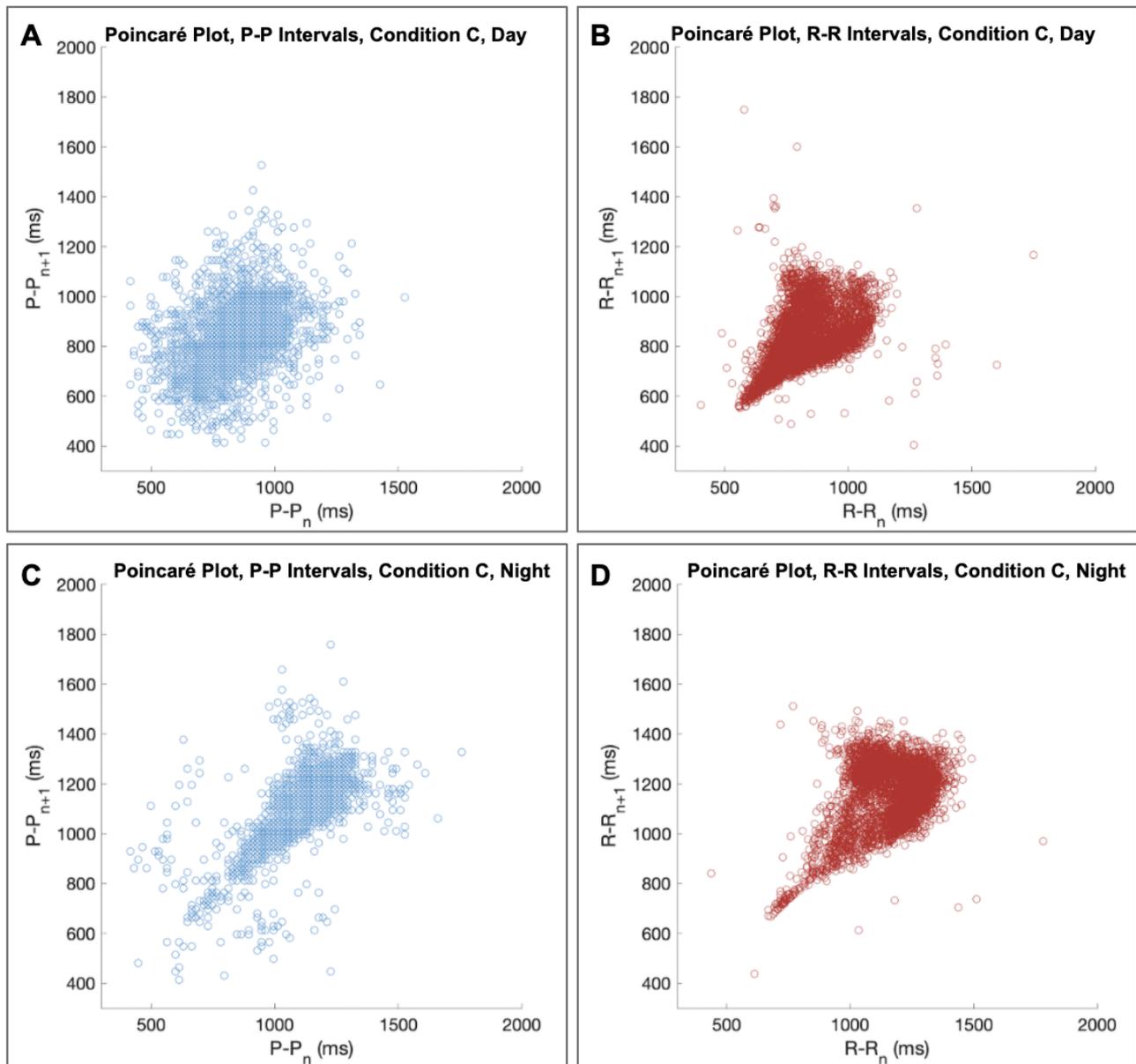
**Figure 5.** Time series of matched time between 2 R peaks in an electrocardiogram and time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals for a single participant under processing condition C during (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.



Poincaré plots for the same participant under condition C are shown in Figure 6. The P-P and R-R plots during the day appear qualitatively different. Although plots of night data demonstrate

more similarities, a greater number of outliers for shorter P-P intervals were observed.

**Figure 6.** Poincaré plots for a single participant under processing condition C for (a) P-P intervals during the day, (b) R-R intervals during the day, (c) P-P intervals at night, and (d) R-R intervals at night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.

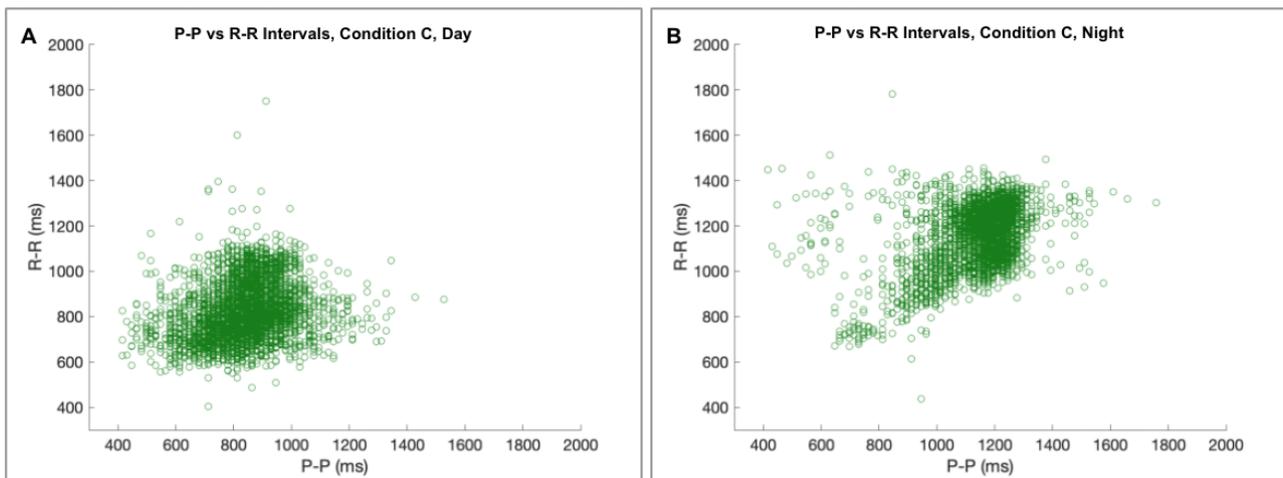


### Comparison of Free-Living Data Collection Conditions (Day vs Night)

Table 2 shows the difference in concurrent validity for night data versus day data under condition C. Closer examination of the data reveals further details. For 8 of 9 participants with a day and night data set, the average  $R^2$  values were higher at

night. The increase in  $R^2$  is highlighted for one participant in Figure 7, where the  $R^2_{\text{day}}=0.26$  and  $R^2_{\text{night}}=0.40$ . The magnitude of  $R^2$  improvements from day to night differed between participants, ranging from  $-0.03$  to  $0.60$  with an average improvement of  $0.22$  (SD  $0.31$ ). Paired  $t$  tests comparing changes in  $R^2$  were significant ( $P=.01$ ).

**Figure 7.** P-P versus R-R intervals for a participant under processing condition C during (A) day and (B) night. P-P: time between 2 P peaks in a photoplethysmogram or peak-to-peak intervals; R-R: time between 2 R peaks in an electrocardiogram.



Night collections were found to have a slight decrease in RMSE, indicated by a mean decrease in RMSE of 24 (SD 45) ms, ranging from  $-89$  ms to  $+40$  ms difference across participants. For the participant highlighted in Figure 7,  $RMSE_{day}=148$  ms and  $RMSE_{night}=138$  ms. Paired  $t$  tests comparing changes in RMSE from day to night approached significance ( $P=.09$ ).

Although night data had more matched samples, an unpaired  $t$  test revealed that the difference between night and day samples was significant ( $P=.03$ ). The mean percent increase in samples from day to night was 138.61% (SD 159). Differences in percentage loss of data owing to filtering (condition C vs condition A) were slightly higher during the day, averaging 13.31% (SD 11.47), versus night, 8.16% (SD 6.05). This difference did not reach significance under the unpaired  $t$  test ( $P=.25$ ).

Tables 2 and 3 demonstrate that many PRV estimates of HRV measures were more accurate at night, with  $|Error|_{avg}$  decreasing or remaining the same for NN, SDNN, RMSSD, SD1, SD2, and LF/HF ratio.  $|Error|_{avg}$  for LF and HF remained approximately the same, whereas  $|Error|_{avg}$  increased for pNN50 at night. Although  $|Error|_{avg}$  generally decreased, paired  $t$  tests revealed more differences between PRV and HRV estimates across participants for night samples than day.

## Discussion

### Principal Findings

This paper examined the accuracy and concurrent validity of PRV measurements from a commercially available PPG sensor against HRV measurements obtained from a portable ECG sensor during unsupervised daytime and nighttime conditions. Accuracy and concurrent validity were examined across different editing methods and day and night collection conditions. In general, concurrent validity and HRV metrics were stronger at night compared with daytime conditions. Although collection during the night was more accurate with a lower mean error, this finding was not generalizable across all participants. Editing to remove outliers was effective in reducing noise, as reflected by the reduced RMSE for conditions B, C, and D. However,

efforts to remove samples affected by motion artifacts using accelerometry (ie, condition D) were not as effective in this study compared with previous studies. The implications of these findings on ambulatory measurement of HRV using a commercially available PPG sensor to indicate health are discussed.

Although PPG sensors have strong mean HR measurement capabilities, the results from this study indicate poorer HRV capabilities. As expected, both ECG and PPG methods demonstrated similar mean R-R values with differences of less than 20 ms, reflecting established capabilities to estimate mean HR [17,20]. Examining beat-to-beat intervals using Bland-Altman plots, the mean error is close to zero (Figure 3). However, the wide variability of both under- and overestimated intervals indicates the presence of error-inducing factors, reflected in lower correlation ( $R^2$ ) and large differences in calculated HRV metrics. Furthermore, Bland-Altman (Figure 3), time series (Figure 5), and Poincaré (Figure 6) plots indicate PPG sensing to trend toward underestimation errors.

The implications of PPG sensing errors on HRV metrics are highlighted in Table 3. pNN50 and LF/HF ratios were particularly sensitive to errors in point-to-point accuracy. PPG-derived estimates of pNN50 were poor, which corroborates previous reports of up to 30% error [12,21]. In addition, LF/HF ratio estimation errors were anticipated to be related to poorer HF estimates during the day arising from larger and more frequent (wrist) motion associated with regular activities of daily living. Across day and night collection conditions, SDNN estimates were similar when comparing ECG and PPG methods. SDNN has been shown to be associated with daytime occupational stress and has been hypothesized to demonstrate the parasympathetic autoregulation of the cardiac system in response to variations in cardiac output [45,46].

### Day Versus Night Collection

When comparing day and night collection conditions, concurrent validity and HRV metrics indicate more accurate HRV estimates at night. Improved concurrent validity at night may be attributed to fewer errors related to ambient light changes at night [47,48], as presumably during sleep, the lighting conditions are

consistently darker. An important distinction between day and night was the larger sample size at night, likely owing to a more consistent Bluetooth stream and reduced noise arising from stationary conditions at night (ie, sleeping, lying down). Although mean  $R^2$  values (condition C: mean 0.34, SD 0.21; condition D: mean 0.34, SD 0.21) were highest during night collections, the range of improvements varied across participants. Conversely, paired  $t$  tests revealed greater differences between PRV and HRV metrics (Table 3) at night. This may be attributed to the larger variability observed during the day, likely associated with a larger set and magnitude of motion during activities of daily living (compared with night). Considering significantly stronger concurrent validity measures (Table 2), coupled with smaller mean differences in HRV measures, we consider PRV estimates to better reflect HRV metrics at night. Although data collected at night may have improved concurrent validity, it is important to note that individuals may not have gone to bed or fallen asleep immediately after beginning the data collection. As a result, metrics taken at night may have captured features of wakefulness, such as shorter N-N intervals. For example, unaccounted for time in bed while remaining awake may have skewed the shape of the Poincaré plots as well as metric SD2.

### Impact of Editing

Simple editing methods to improve PPG signals were examined in this study. PPG recordings are known to be affected by motion artifacts, contact force, posture, and ambient temperature [11]. Owing to the free-living nature of the study, the latter factors were not controlled. By adopting methods established by Morelli et al [28], RMSE improved by removing physiologically implausible intervals (condition B) and concurrent validity improved by deleting areas with rapid changes (condition C) at night. However, screening for motion artifacts using accelerometry signals (condition D) was ineffective at improving PPG-derived signals and HRV estimates. This is consistent with the findings by Baek and Shin [30], who were also unsuccessful in obtaining accurate long-term free-living recordings of wrist PPG using a custom device, even when performing deletions based on acceleration and P-P intervals differing by more than 15%. These findings, along with those by Georgiou et al [20], suggest that under unrestricted conditions, PRV is a poor estimator of HRV. Although other studies have looked to improve HRV estimation using alternative editing and correction methods [39,44,49], an exhaustive investigation of correction methods is beyond the scope of this study.

Our finding of relatively ineffective use of motion artifact compensation suggests that other factors affect PPG signals. For example, changes in respiration and peripheral vascular factors (ie, vascular volume, vasomotor activity, and vasoconstrictor waves) are known to affect the AC and DC frequency components of the PPG waveform [50]. In particular, the effect of peripheral vascular factors affects pulse transit time (PTT), or the time delay required for blood to travel between the heart and peripheral tissues [17]. Considering the range of daily activities (eg, body position changes [51], stress, and physical activity) lead to fluctuations in blood pressure [52], an

assumed constant PTT is a likely source of error in estimating PRV parameters.

### Limitations

The primary limitations of this study were the sample population and technical limitations of the devices. In this study, a convenience sample of 18- to 65-year-old participants with no known cardiac history participated. Although those with known cardiac conditions were excluded, the presence of underlying vascular disease in our cohort is unknown. As such, the findings may not be applicable to target disease populations. The impact of vascular conditions, such as atherosclerosis and cholesterol deposits in the arterial walls leading to decreased vessel compliance, which have been shown to alter pulse waveform from the classic triphasic pattern to mono- or biphasic patterns [53], remains to be examined. Although the number of participants was relatively small ( $n=10$ ), the large number of within-participant samples ( $>1500$  matched interval points per participant) and analyses supports the overall questions regarding sensor comparisons to estimate HRV.

The devices used in this study were limited in several ways. Both Shimmer3 and MB2 devices logged using separate device clocks, with potential for drift (approximately 1-2 s) over the course of a single trial. The devices were synchronized using an external mechanical stimulus (ie, 3 taps in 2 orientations) and by applying a data-driven delay estimate (ie, cross-correlation). Although these procedures have been used in previous studies with good results coupled with qualitative and quantitative observation of synchronized signals, the potential for dropped samples or desynchronization exists. The publicly available documentation for MB2 offers little to no insight into R-R interval processing or adjustment for when faced with motion artifacts and is no longer commercially available at the time of writing. Of note, signal drops were observed sporadically, including (1) large amplitude arm movements and (2) when MB2 was out of Bluetooth range from the smartphone for long periods ( $>10$  min). We interpret these signal drops as obvious situations where motion artifacts and wireless communication are severely challenged with little to no impact on our findings. Furthermore, the resolution of RR intervals reported by MB2 was 10 ms, limiting accuracy similar to quantization error (ie, round-off errors). Given the large number of samples, resolution limitations are unlikely to affect mean values (eg, mean RR) but may increase variability (eg, RMSSD) estimates. However, the observed underestimation is unlikely to arise from quantization errors and are interpreted as systematic errors associated with the sensing method.

### Implications for Future Work

Wearable technologies are becoming more sophisticated with commercially available products capable of providing consumers access to information previously limited to clinical settings, including HRV and ECG data to identify arrhythmias [54]. With this in mind, it is important to understand when and if the data can be considered valid and reliable. This study provides evidence that the relationship between PRV and HRV varies throughout the day, likely attributable to dynamic changes in the peripheral vasculature. The study findings suggest that PPG-derived measures of HRV are reasonable under particular

conditions (ie, at night), wherein this relationship is relatively stable for some HRV metrics (ie, SDNN and Poincaré axes). A deeper examination of factors modifying HRV estimation, particularly vascular factors, is yet to be conducted. As stated previously, our conclusions are drawn primarily from the HR and acceleration data. To further study these windows of high correlation in the future, other variables such as body temperature, cortisol levels, or a cognitive assessment of the participant's mental state may be beneficial.

In future, examining more editing, correction, and interpolation techniques for interbeat intervals may enhance the interpretability and quality of the P-P intervals obtained from commercially available wearables [44,55-57]. This study found that published movement artifact reduction techniques did not significantly improve the quality of our data. As wearable technologies continue to become more advanced, future studies in this field would benefit from the use of improved hardware and more robust sensors. For example, PulseOn and Apple Watch, both commercially available wearable devices, use different strategies to improve the quality of their signals. PulseOn uses multiwavelength PPG to reduce the sensitivity to movement artifacts and ambient light disturbances [47,48], demonstrating 99.57% accuracy during sleep [47]. Considering the lack of peripheral vascular indicators to account for changes

in PTT, the Apple Watch approach of directly acquiring R-R intervals using built-in or peripheral ECG sensors (Kardia Band, Alivecor) [58] is justifiable.

## Conclusions

The objective of this study was to assess the validity of PRV measurements taken from a PPG sensor by comparing it with the HRV measurements taken from a portable ECG while individuals were engaged in activities of daily living and during sleep. Although PPG sensors demonstrated greater validity at night, overall concurrent validity was poor. HRV metrics pNN50 and LF/HF ratio were especially sensitive to errors in point-to-point accuracy. Increased editing via deletion improved the RMSE but had a small impact on  $R^2$ . In comparing editing and deletion methods, screening for motion artifacts using accelerometry signals to remove error-prone signals was largely ineffective in improving HRV estimates. The best results were obtained under condition C (moving average method) at night, with the highest mean  $R^2$  values. Overall, the findings from this study suggest that PRV is a poor surrogate of HRV under free-living conditions. Findings from this study indicate that advances in hardware and wearable technologies, such as multiwavelength PPG sensors, are warranted to unleash the potential of PRV to serve as a proxy measure for HRV.

## Acknowledgments

This study was supported by the National Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (RGPIN-2015-05317).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Comparison of mean heart rate variability and pulse rate variability metrics under processing condition D. [\[DOCX File, 15 KB-Multimedia Appendix 1\]](#)

## References

- Heart rate variability. In: *Advances in Cardiac Signal Processing*. New York, USA: Springer; 2007.
- Bianchi AM. Signal Processing and Feature Extraction for Sleep Evaluation in Wearable Devices. In: *International Conference of the IEEE Engineering in Medicine and Biology Society*. 2006 Presented at: IEMBS'06; August 30-September 4, 2006; New York, NY, USA. [doi: [10.1109/iembs.2006.260547](https://doi.org/10.1109/iembs.2006.260547)]
- Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: <http://los40.com>, <http://los40principales.com>, <http://cadena100.es>, <http://europafm.es> and <http://kissfm.es>. *J Int Stud* 2009 Feb 1;6(1):5-20 [FREE Full text] [doi: [10.1386/rajo.6.1.5\\_4](https://doi.org/10.1386/rajo.6.1.5_4)]
- Chuang C, Ye J, Lin W, Lee K, Tai Y. Photoplethysmography variability as an alternative approach to obtain heart rate variability information in chronic pain patient. *J Clin Monit Comput* 2015 Dec;29(6):801-806. [doi: [10.1007/s10877-015-9669-8](https://doi.org/10.1007/s10877-015-9669-8)] [Medline: [25708672](https://pubmed.ncbi.nlm.nih.gov/25708672/)]
- Tan G, Dao TK, Farmer L, Sutherland RJ, Gevirtz R. Heart rate variability (HRV) and posttraumatic stress disorder (PTSD): a pilot study. *Appl Psychophysiol Biofeedback* 2011 Mar;36(1):27-35. [doi: [10.1007/s10484-010-9141-y](https://doi.org/10.1007/s10484-010-9141-y)] [Medline: [20680439](https://pubmed.ncbi.nlm.nih.gov/20680439/)]
- Liao D, Carnethon M, Evans GW, Cascio WE, Heiss G. Lower heart rate variability is associated with the development of coronary heart disease in individuals with diabetes: the atherosclerosis risk in communities (ARIC) study. *Diabetes* 2002 Dec;51(12):3524-3531 [FREE Full text] [doi: [10.2337/diabetes.51.12.3524](https://doi.org/10.2337/diabetes.51.12.3524)] [Medline: [12453910](https://pubmed.ncbi.nlm.nih.gov/12453910/)]
- Reed M, Robertson C, Addison P. Heart rate variability measurements and the prediction of ventricular arrhythmias. *QJM* 2005 Feb;98(2):87-95. [doi: [10.1093/qjmed/hci018](https://doi.org/10.1093/qjmed/hci018)] [Medline: [15671474](https://pubmed.ncbi.nlm.nih.gov/15671474/)]

8. Martin T, Jovanov E, Raskovic D. Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device. In: Fourth International Symposium on Wearable Computers. 2000 Presented at: ISCW'00; October 16-17, 2000; Atlanta, GA, USA. [doi: [10.1109/iswc.2000.888463](https://doi.org/10.1109/iswc.2000.888463)]
9. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. *J Med Internet Res* 2018 Mar 22;20(3):e110 [FREE Full text] [doi: [10.2196/jmir.9157](https://doi.org/10.2196/jmir.9157)] [Medline: [29567635](https://pubmed.ncbi.nlm.nih.gov/29567635/)]
10. Chou Y, Zhang R, Feng Y, Lu M, Lu Z, Xu B. A real-time analysis method for pulse rate variability based on improved basic scale entropy. *J Healthc Eng* 2017;2017:7406896 [FREE Full text] [doi: [10.1155/2017/7406896](https://doi.org/10.1155/2017/7406896)] [Medline: [29065639](https://pubmed.ncbi.nlm.nih.gov/29065639/)]
11. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. *Electronics* 2014 Apr 23;3(2):282-302. [doi: [10.3390/electronics3020282](https://doi.org/10.3390/electronics3020282)]
12. Jeyhani V, Mahdiani S, Peltokangas M, Vehkaoja A. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:5952-5955. [doi: [10.1109/EMBC.2015.7319747](https://doi.org/10.1109/EMBC.2015.7319747)] [Medline: [26737647](https://pubmed.ncbi.nlm.nih.gov/26737647/)]
13. -. Comparison of Heart Rate Variability from PPG with That from ECG. In: The International Conference on Health Informatics. 2014 Presented at: CHI'14; September 15-17, 2014; Verona, Italy. [doi: [10.1007/978-3-319-03005-0\\_54](https://doi.org/10.1007/978-3-319-03005-0_54)]
14. Bolanos M, Nazeran H, Haltiwanger E. Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. *Conf Proc IEEE Eng Med Biol Soc* 2006;2006:4289-4294. [doi: [10.1109/IEMBS.2006.260607](https://doi.org/10.1109/IEMBS.2006.260607)] [Medline: [17946618](https://pubmed.ncbi.nlm.nih.gov/17946618/)]
15. Lu G, Yang F, Taylor JA, Stein JF. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *J Med Eng Technol* 2009;33(8):634-641. [doi: [10.3109/03091900903150998](https://doi.org/10.3109/03091900903150998)] [Medline: [19848857](https://pubmed.ncbi.nlm.nih.gov/19848857/)]
16. Russoniello CV, Pougatchev V, Zhirnov E, Mahar MT. A measurement of electrocardiography and photoplethysmography in obese children. *Appl Psychophysiol Biofeedback* 2010 Sep;35(3):257-259. [doi: [10.1007/s10484-010-9136-8](https://doi.org/10.1007/s10484-010-9136-8)] [Medline: [20552266](https://pubmed.ncbi.nlm.nih.gov/20552266/)]
17. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *J Med Eng Technol* 2008;32(6):479-484. [doi: [10.1080/03091900701781317](https://doi.org/10.1080/03091900701781317)] [Medline: [18663635](https://pubmed.ncbi.nlm.nih.gov/18663635/)]
18. Cropley M, Plans D, Morelli D, Sütterlin S, Inceoglu I, Thomas G, et al. The association between work-related rumination and heart rate variability: a field study. *Front Hum Neurosci* 2017;11:27 [FREE Full text] [doi: [10.3389/fnhum.2017.00027](https://doi.org/10.3389/fnhum.2017.00027)] [Medline: [28197087](https://pubmed.ncbi.nlm.nih.gov/28197087/)]
19. Amoedo A, Martinez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: <http://los40.com>, <http://los40principales.com>, <http://cadena100.es>, <http://europafm.es> and <http://kissfm.es>. *radio journal: international studies* in 2009 Feb 01;6(1):5-20 [FREE Full text] [doi: [10.1386/rajo.6.1.5\\_4](https://doi.org/10.1386/rajo.6.1.5_4)]
20. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. *Folia Med (Plovdiv)* 2018 Mar 1;60(1):7-20. [doi: [10.2478/foimed-2018-0012](https://doi.org/10.2478/foimed-2018-0012)] [Medline: [29668452](https://pubmed.ncbi.nlm.nih.gov/29668452/)]
21. 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 5;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/)]
22. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiol Meas* 2010 Sep;31(9):1271-1290. [doi: [10.1088/0967-3334/31/9/015](https://doi.org/10.1088/0967-3334/31/9/015)] [Medline: [20702919](https://pubmed.ncbi.nlm.nih.gov/20702919/)]
23. Salehizadeh S, Dao D, Bolkhovsky J, Cho C, Mendelson Y, Chon K. A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor. *Sensors (Basel)* 2015 Dec 23;16(1):- [FREE Full text] [doi: [10.3390/s16010010](https://doi.org/10.3390/s16010010)] [Medline: [26703618](https://pubmed.ncbi.nlm.nih.gov/26703618/)]
24. Jarchi D, Casson A. Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion. *Data* 2016 Dec 24;2(1):1 [FREE Full text] [doi: [10.3390/data2010001](https://doi.org/10.3390/data2010001)]
25. Ram MR, Madhav KV, Krishna EH, Komalla NR, Reddy KA. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Trans Instrum Meas* 2012 May;61(5):1445-1457. [doi: [10.1109/tim.2011.2175832](https://doi.org/10.1109/tim.2011.2175832)]
26. Lee C, Zhang Y. Reduction of Motion Artifacts From Photoplethysmographic Recordings Using a Wavelet Denoising Approach. In: Asian-Pacific Conference on Biomedical Engineering. 2003 Presented at: Kyoto, Japan; October 20-23, 2003; Kyoto, Japan. [doi: [10.1109/apbme.2003.1302650](https://doi.org/10.1109/apbme.2003.1302650)]
27. Kim B, Yoo S. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans Biomed Eng* 2006 Mar;53(3):566-568. [doi: [10.1109/tbme.2005.869784](https://doi.org/10.1109/tbme.2005.869784)]
28. Morelli D, Bartoloni L, Colombo M, Plans D, Clifton DA. Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. *Healthc Technol Lett* 2018 Apr;5(2):59-64 [FREE Full text] [doi: [10.1049/htl.2017.0039](https://doi.org/10.1049/htl.2017.0039)] [Medline: [29750114](https://pubmed.ncbi.nlm.nih.gov/29750114/)]

29. Kos M, Khaghani-Far I, Gordon CM, Pavel M, Jimison HB. Can accelerometry data improve estimates of heart rate variability from wrist pulse PPG sensors? *Conf Proc IEEE Eng Med Biol Soc* 2017 Jul;2017:1587-1590 [FREE Full text] [doi: [10.1109/EMBC.2017.8037141](https://doi.org/10.1109/EMBC.2017.8037141)] [Medline: [29060185](https://pubmed.ncbi.nlm.nih.gov/29060185/)]
30. Baek HJ, Shin J. Effect of missing inter-beat interval data on heart rate variability analysis using wrist-worn wearables. *J Med Syst* 2017 Aug 15;41(10):147. [doi: [10.1007/s10916-017-0796-2](https://doi.org/10.1007/s10916-017-0796-2)] [Medline: [28812280](https://pubmed.ncbi.nlm.nih.gov/28812280/)]
31. *The ECG Manual: An Evidence-Based Approach*. New York, USA: Springer; 2018.
32. R Wave Detection in the ECG. MathWorks. URL: <https://www.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html> [accessed 2019-08-09]
33. Li C, Zheng C, Tai C. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans Biomed Eng* 1995 Jan;42(1):21-28. [doi: [10.1109/10.362922](https://doi.org/10.1109/10.362922)] [Medline: [7851927](https://pubmed.ncbi.nlm.nih.gov/7851927/)]
34. Manurmath JC, Raveendra M. CMATLAB based ECG signal classification. *Int J SciEng Technol Res* 1946:1950.
35. Agelink MW, Malessa R, Baumann B, Majewski T, Akila F, Zeit T, et al. Standardized tests of heart rate variability: normal ranges obtained from 309 healthy humans, and effects of age, gender, and heart rate. *Clin Auton Res* 2001 Apr;11(2):99-108. [doi: [10.1007/BF02322053](https://doi.org/10.1007/BF02322053)] [Medline: [11570610](https://pubmed.ncbi.nlm.nih.gov/11570610/)]
36. Murray MP, Spurr GB, Sepic SB, Gardner GM, Mollinger LA. Treadmill vs floor walking: kinematics, electromyogram, and heart rate. *J Appl Physiol* (1985) 1985 Jul;59(1):87-91. [doi: [10.1152/jap.1985.59.1.87](https://doi.org/10.1152/jap.1985.59.1.87)] [Medline: [4030579](https://pubmed.ncbi.nlm.nih.gov/4030579/)]
37. Snyder F, Hobson JA, Morrison DF, Goldfrank F. Changes in respiration, heart rate, and systolic blood pressure in human sleep. *J Appl Physiol* 1964 May 1;19(3):417-422. [doi: [10.1152/jap.1964.19.3.417](https://doi.org/10.1152/jap.1964.19.3.417)]
38. 38 L. Reinerman-Jones, J. Harris, and A. Watson, "Considerations for Using Fitness Trackers in Psychophysiology Research," *Human Interface and the Management of Information: Information, Knowledge and Interaction Design*. pp. 598-606 2017:-. [doi: [10.1007/978-3-319-58521-5\\_47](https://doi.org/10.1007/978-3-319-58521-5_47)]
39. Giles DA, Draper N. Heart rate variability during exercise: a comparison of artefact correction methods. *J Strength Cond Res* 2018 Mar;32(3):726-735. [doi: [10.1519/JSC.0000000000001800](https://doi.org/10.1519/JSC.0000000000001800)] [Medline: [29466273](https://pubmed.ncbi.nlm.nih.gov/29466273/)]
40. Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: <http://los40.com>, <http://los40principales.com>, <http://cadena100.es>, <http://europafm.es> and <http://kissfm.es>. *J Int Stud* 2009 Feb 1;6(1):5-20 [FREE Full text] [doi: [10.1386/rajo.6.1.5\\_4](https://doi.org/10.1386/rajo.6.1.5_4)]
41. Li K, Rüdiger H, Ziemssen T. Spectral analysis of heart rate variability: time window matters. *Front Neurol* 2019;10:545 [FREE Full text] [doi: [10.3389/fneur.2019.00545](https://doi.org/10.3389/fneur.2019.00545)] [Medline: [31191437](https://pubmed.ncbi.nlm.nih.gov/31191437/)]
42. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med* 2018 Sep;18(3):91-93 [FREE Full text] [doi: [10.1016/j.tjem.2018.08.001](https://doi.org/10.1016/j.tjem.2018.08.001)] [Medline: [30191186](https://pubmed.ncbi.nlm.nih.gov/30191186/)]
43. Giavarina D. Understanding bland altman analysis. *Biochem Med (Zagreb)* 2015;25(2):141-151 [FREE Full text] [doi: [10.11613/BM.2015.015](https://doi.org/10.11613/BM.2015.015)] [Medline: [26110027](https://pubmed.ncbi.nlm.nih.gov/26110027/)]
44. Morelli D, Rossi A, Cairo M, Clifton DA. Analysis of the impact of interpolation methods of missing RR-intervals caused by motion artifacts on HRV features estimations. *Sensors (Basel)* 2019 Jul 18;19(14):- [FREE Full text] [doi: [10.3390/s19143163](https://doi.org/10.3390/s19143163)] [Medline: [31323850](https://pubmed.ncbi.nlm.nih.gov/31323850/)]
45. Borchini R, Bertù L, Ferrario MM, Veronesi G, Bonzini M, Dorso M, et al. Prolonged job strain reduces time-domain heart rate variability on both working and resting days among cardiovascular-susceptible nurses. *Int J Occup Med Environ Health* 2015;28(1):42-51 [FREE Full text] [doi: [10.2478/s13382-014-0289-1](https://doi.org/10.2478/s13382-014-0289-1)] [Medline: [26159946](https://pubmed.ncbi.nlm.nih.gov/26159946/)]
46. Billman GE. The effect of heart rate on the heart rate variability response to autonomic interventions. *Front Physiol* 2013;4:222 [FREE Full text] [doi: [10.3389/fphys.2013.00222](https://doi.org/10.3389/fphys.2013.00222)] [Medline: [23986716](https://pubmed.ncbi.nlm.nih.gov/23986716/)]
47. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. *Conf Proc IEEE Eng Med Biol Soc* 2015 Aug;2015:8099-8102. [doi: [10.1109/EMBC.2015.7320273](https://doi.org/10.1109/EMBC.2015.7320273)] [Medline: [26738173](https://pubmed.ncbi.nlm.nih.gov/26738173/)]
48. Renevey PH, Sola J, Theurillat P, Bertschi M, Krauss J, Andries D, et al. Validation of a wrist monitor for accurate estimation of RR intervals during sleep. *Conf Proc IEEE Eng Med Biol Soc* 2013;2013:5493-5496. [doi: [10.1109/EMBC.2013.6610793](https://doi.org/10.1109/EMBC.2013.6610793)] [Medline: [24110980](https://pubmed.ncbi.nlm.nih.gov/24110980/)]
49. Lee J, Kim J, Shin M. Correlation analysis between electrocardiography (ECG) and photoplethysmogram (PPG) data for driver's drowsiness detection using noise replacement method. *Procedia Computer Science* 2017;116:421-426 [FREE Full text] [doi: [10.1016/j.procs.2017.10.083](https://doi.org/10.1016/j.procs.2017.10.083)]
50. Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. *J Med Syst* 2011 Oct;35(5):969-976. [doi: [10.1007/s10916-010-9505-0](https://doi.org/10.1007/s10916-010-9505-0)] [Medline: [20703691](https://pubmed.ncbi.nlm.nih.gov/20703691/)]
51. Olufsen MS, Ottesen JT, Tran HT, Ellwein LM, Lipsitz LA, Novak V. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. *J Appl Physiol* (1985) 2005 Oct;99(4):1523-1537 [FREE Full text] [doi: [10.1152/jap.1985.99.4.1523](https://doi.org/10.1152/jap.1985.99.4.1523)] [Medline: [15860687](https://pubmed.ncbi.nlm.nih.gov/15860687/)]
52. Drinnan MJ, Allen J, Murray A. Relation between heart rate and pulse transit time during paced respiration. *Physiol Meas* 2001 Aug;22(3):425-432. [doi: [10.1088/0967-3334/22/3/301](https://doi.org/10.1088/0967-3334/22/3/301)] [Medline: [11556663](https://pubmed.ncbi.nlm.nih.gov/11556663/)]
53. Azzopardi YM, Gatt A, Chockalingam N, Formosa C. Agreement of clinical tests for the diagnosis of peripheral arterial disease. *Prim Care Diabetes* 2019 Feb;13(1):82-86. [doi: [10.1016/j.pcd.2018.08.005](https://doi.org/10.1016/j.pcd.2018.08.005)] [Medline: [30201222](https://pubmed.ncbi.nlm.nih.gov/30201222/)]

54. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the apple heart study. *Am Heart J* 2019 Jan;207:66-75 [FREE Full text] [doi: [10.1016/j.ahj.2018.09.002](https://doi.org/10.1016/j.ahj.2018.09.002)] [Medline: [30392584](https://pubmed.ncbi.nlm.nih.gov/30392584/)]
55. Lang M. Automatic near real-time outlier detection and correction in cardiac interbeat interval series for heart rate variability analysis: singular spectrum analysis-based approach. *JMIR Biomed Eng* 2019 Jan 30;4(1):e10740 [FREE Full text] [doi: [10.2196/10740](https://doi.org/10.2196/10740)]
56. Citi L, Brown EN, Barbieri R. A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. *IEEE Trans Biomed Eng* 2012 Oct;59(10):2828-2837 [FREE Full text] [doi: [10.1109/TBME.2012.2211356](https://doi.org/10.1109/TBME.2012.2211356)] [Medline: [22875239](https://pubmed.ncbi.nlm.nih.gov/22875239/)]
57. Tarvainen MP, Niskanen J, Lippinen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. *Comput Methods Programs Biomed* 2014;113(1):210-220. [doi: [10.1016/j.cmpb.2013.07.024](https://doi.org/10.1016/j.cmpb.2013.07.024)] [Medline: [24054542](https://pubmed.ncbi.nlm.nih.gov/24054542/)]
58. Bumgarner JM. Smartwatch algorithm for automated detection of atrial fibrillation. *J Am Coll Cardiol* 2018;71(21):-.

## Abbreviations

**ECG:** electrocardiogram  
**HF:** high-frequency power (0.15-0.40 Hz)  
**HR:** heart rate  
**HRV:** heart rate variability  
**LA:** left arm  
**LF:** low-frequency power (0.04-0.15 Hz)  
**LL:** left leg  
**MB2:** Microsoft Band 2  
**N-N:** N-N intervals (either R-R or P-P intervals)  
**pNN50:** percent of subsequent differences more than 50 ms  
**P-P:** time between 2 P peaks in a PPG or peak-to-peak intervals  
**PPG:** photoplethysmography  
**PR:** pulse rate  
**PRV:** pulse rate variability  
**PTT:** pulse transit time  
**R2:** Pearson correlation coefficient  
**RA:** right arm  
**RMSE:** root mean square error  
**RMSSD:** root mean square of subsequent differences  
**R-R:** time between 2 R peaks in an ECG  
**SD1:** standard deviation of short (x=y) Poincaré plot axis  
**SD2:** standard deviation of long (orthogonal to x=y) Poincaré plot axis  
**SDNN:** standard deviation of all N-N intervals, also known as RRSD  
**Wt:** mean change in triaxial acceleration

*Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by A Vehkaoja, M Lang, C Boodoo; comments to author 10.02.20; revised version received 31.05.20; accepted 26.07.20; published 03.11.20*

*Please cite as:*

Lam E, Aratia S, Wang J, Tung J

*Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study*

*JMIR Biomed Eng* 2020;5(1):e17355

URL: <http://biomedeng.jmir.org/2020/1/e17355/>

doi: [10.2196/17355](https://doi.org/10.2196/17355)

PMID:

©Emily Lam, Shahrose Aratia, Julian Wang, James Tung. Originally published in *JMIR Biomedical Engineering* (<http://biomedeng.jmir.org>), 03.11.2020. 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 Biomedical Engineering*, is properly cited. The complete bibliographic information, a link to the original publication on <http://biomedeng.jmir.org/>, as well as this copyright and license information must be included.