Med Bio Eng Comput (2006) 44:747–756
DOI 10.1007/s11517-006-0097-2

ORIGINAL ARTICLE

# Reliability and accuracy of heart rate variability metrics versus ECG segment duration

**James McNames · Mateo Aboy**

Received: 28 March 2006 / Accepted: 21 July 2006 / Published online: 22 August 2006


**Abstract** Despite the exponential growth in heart rate variability (HRV) research, the reproducibility and reliability of HRV metrics continues to be debated. We estimated the reliability of 11 metrics calculated from 5 min records. We also compared the accuracy of the HRV metrics calculated from ECG records spanning 10 s to 10 min as compared with the metrics calculated from 5 min records. The mean heart rate was more reproducible and could be more accurately estimated from very short segments (<1 min) than any of the other HRV metrics. HRV metrics that effectively highpass filter the R–R interval series were more reliable than the other metrics and could be more accurately estimated from very short segments. This indicates that most of the HRV is caused by drift and nonstationary effects. Metrics that are sensitive to low frequency components of HRV have poor repeatability and cannot be estimated accurately from short segments (<10 min).

**Keywords** Heart rate variability (HRV) · Interbeat intervals (IBI) · Intraclass correlation coefficient (ICC) · Reliability · Reproducibility · Repeatability · Consistency

J. McNames
Electrical and Computer Engineering,
Portland State University, P.O. Box 751,
Portland, OR 97207-0751, USA
e-mail: mcnames@pdx.edu

M. Aboy (✉)
Electronics Engineering Technology,
Oregon Institute of Technology,
18640 NW Walker Rd,
Beaverton, OR 97006, USA
e-mail: mateoaboy@ieee.org

## 1 Introduction

Heart rate variability (HRV) has received a tremendous amount of attention since the seminal work of Akselrod et al. [1]. Established clinical applications of HRV include risk assessment of patients after myocardial infarction and early diagnosis of diabetic autonomic neuropathy [24, 57]. The physiologic mechanisms underlying HRV continue to be investigated [7, 26, 36, 52].

A task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology published standards of measurement, interpretation and use of HRV in 1996 [57]. The task force specified many different HRV metrics for both short-term records (5 min) and long-term records (24 h). Although many other measures of HRV have been proposed and investigated, those specified by the task force have been the most widely applied.

Despite the exponential growth in HRV research and the number of studies published addressing the reliability, repeatability and reproducibility of HRV metrics, this topic continues to be debated. Some studies concluded that their examined HRV metrics were reliable [2, 5, 8, 13, 14, 18, 19, 25, 29, 31, 34, 35, 39, 41, 50, 55, 58, 59, 60, 61], while others found the reliability to be moderate or low [6, 23, 33, 51, 56, 62]. This discrepancy of results may be due, in part, to the fact that these studies examined distinct subject populations, and their conclusions on HRV reliability apply only to the specific HRV metrics studied.

In this paper, we address how the accuracy of time- and frequency-domain HRV metrics varies with the record duration and estimate the reliability of these metrics calculated from 5 min records. In order to address the question of HRV reliability, we studied the reliability of the most widely used HRV metrics in three different databases. We used publicly available databases to ensure the reproducibility of our results. Additionally, we propose a new methodology and visualization technique which can be used to assess the segment length required for a given degree of accuracy.

## 2 HRV metrics

We studied four time-domain and six frequency-domain metrics, recommended by the task force for short-term ECG records. We also studied approximate entropy as a measure of the complexity of HRV. We denote each of the $N$ beat times as $t(n)$ for $n\in\{1,...,N\}$. We denote the interval between beats as $\delta(n) = t(n) - t(n-1)$. As recommended by the task force, we define the time of occurrence of each interval $\delta(n)$ as $t(n)$.

### 2.1 Time domain metrics

#### *2.1.1 Standard deviation of the normal-to-normal*

We calculated the standard deviation of the normal-to-normal (NN) intervals (SDNN) as

$$\mathrm{SDNN} = \sqrt{\frac{1}{N-2}\sum_{n=2}^{N}\left(\delta(n)-\bar{\delta}\right)^2} \tag{1}$$

where $\bar{\delta}$ is the average NN interval,

$$\bar{\delta} = \frac{1}{N-1}\sum_{n=2}^{N}\delta(n). \tag{2}$$

The scaling factor is $N-2$ because there are $N-1$ intervals in the record and one degree of freedom is used to estimate the mean NN interval.

#### *2.1.2 HRV triangular index*

The HRV triangular index (HRVTI) is a measure of the shape of the NN interval distribution. Generally, uniform distributions representing large variability have large values and distributions with single large peaks have small values. The metric is defined in terms of a histogram of the NN intervals. Here, we represent the number of intervals in the $i$th bin centered at $t_i$ as $b(t_i)$. HRVTI is then defined as

$$\mathrm{HRVTI} = \frac{\sum_{i=1}^{N_b} b(t_i)}{\max_i b(t_i)} = \frac{N-1}{\max_i b(t_i)} \tag{3}$$

where $N_b$ is the number of bins. We used a constant bin width of $1/(f_s) = 8$ ms. The task force did not specify how to align the bins. We chose to locate the bin centers at integer multiples of $1/f_s$ so that each bin only contained equal intervals.

#### *2.1.3 Root mean square of successive NN interval differences*

We calculated the root mean square of successive NN interval differences (RMSSD) as

$$\mathrm{RMSSD} = \sqrt{\frac{1}{N-2}\sum_{n=3}^{N}[\delta(n)-\delta(n-1)]^2}. \tag{4}$$

#### *2.1.4 Standard deviation of successive NN interval differences*

We calculated the standard deviation of successive NN interval differences (SDSD) as

$$\mathrm{SDSD} = \sqrt{\frac{1}{N-3}\sum_{n=3}^{N}\left[\delta(n)-\delta(n-1)-\bar{\delta}_\delta\right]^2} \tag{5}$$

where $\bar{\delta}_\delta$ is given by

$$\bar{\delta}_\delta = \frac{1}{N-2}\sum_{n=3}^{N}\delta(n)-\delta(n-1). \tag{6}$$

In most cases $\bar{\delta}_\delta$ is nearly zero and there is little difference between RMSSD and SDSD.

### 2.2 Frequency domain metrics

All frequency domain HRV metrics are based on the estimated power spectral density (PSD) of the NN intervals. Although the task force gave specific definitions of the metrics, it did not specify how to estimate the PSD. There are many methods of estimating PSD and each generates different HRV metric values. If the NN interval series is not well behaved or too short, as in our case, these differences can be substantial. In this section we give a complete description of our PSD estimator, as required by the task force.

We smoothed and uniformly resampled the NN intervals at a rate of 3 Hz with a kernel smoother,

$$d(k) = \frac{\sum_{n=1}^{N} \delta(n) b\left(\frac{|kT_s - t(n)|}{\sigma_b}\right)}{\sum_{n=1}^{N} b\left(\frac{|kT_s - t(n)|}{\sigma_b}\right)} \tag{7}$$

where $T_s = 1/3$ s is the resample interval, $\sigma_b$ is the kernel width, and $b(\cdot)$ is a clipped and scaled Gaussian kernel function

$$\begin{cases} \exp(-u^2/2) & -5 \le u \le 5 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We generated the uniformly resampled signal $d(k)$ over the duration of the ECG record. The kernel smoother also acts like a lowpass filter with a cutoff frequency determined by $\sigma_b$. Large values of $\sigma_b$ reduce aliasing but can bias HRV metrics that rely on accurate estimates of the PSD at higher frequencies. We used $\sigma_b = 0.25$ s.

We applied Welch's nonparametric method of Periodogram averaging to estimate the PSD of $d(k)$. This estimate is calculated as the average of Periodograms calculated from overlapping segments of $d(k)$,

$$\hat{R}(e^{j\omega}) = \frac{1}{KL} \sum_{i=0}^{K-1} \left| \sum_{k=0}^{L-1} \left(d(k+iL) - \bar{d}\right) w(k) e^{-j\omega k} \right|^2 \tag{9}$$

where $K$ is the number of segments, $L$ is the number of samples in each segment, $\bar{d}$ is the sample average of $d(k)$ over the full record, and $w(k)$ is a window function that determines the tradeoff between main lobe width and sideband leakage. We used a Blackman window with a length of 20 s or the duration of the record, whichever is shorter. We allowed a 50% overlap between segments. Each segment was padded with zeros to a total length of 4,096 samples to minimize the error in estimating the signal power over specified frequency ranges with Riemann sums.

#### 2.2.1 Low frequency

The low-frequency (LF) power was calculated as the total signal power in the frequency range of 0.04–0.15 Hz.

#### 2.2.2 High frequency

The high-frequency (HF) power was calculated as the total signal power in the frequency range of 0.15–0.40 Hz.

#### 2.2.3 LF norm

The low-frequency normalized (LF norm) power was calculated as

$$\text{LF norm} = 100 \times \frac{\text{LF}}{\text{TP} - \text{VLF}} \tag{10}$$

where TP is the total signal power is defined below and VLF is defined as the total signal power at frequencies less than 0.04 Hz.

#### 2.2.4 HF norm

The high frequency normalized (HF Norm) power was calculated

$$\text{HF norm} = 100 \times \frac{\text{HF}}{\text{TP} - \text{VLF}}. \tag{11}$$

#### 2.2.5 Low frequency/high frequency

The low frequency–high frequency ratio (LF/HF) was calculated as LF/HF.

#### 2.2.6 Total power

The total power (TP) was calculated as the integral of the PSD estimate over the full frequency range of 0.0–1.5 Hz. This is approximately equal to the variance of $d(k)$.

### 2.3 Complexity metrics

#### 2.3.1 Approximate entropy

Approximate entropy is one of the nonlinear metrics which has been used to analyze the *R–R* intervals in order to estimate reductions of complexity associated with specific pathologies [32, 42]. Approximate entropy was introduced as a quantification of regularity in sequences and time series data, initially motivated by applications to relatively short, noisy data sets. Mathematically it is part of a general development of approximating Markov Chains to a process [44]. It provides a finite sequence formulation of randomness, via proximity to maximal irregularity [47, 49]. A statistical evaluation of approximate entropy (ApEn) is available in [46]. ApEn is scale invariant and model independent, evaluates both dominant and subordinate patterns in data, and discriminates series for which clear feature recognition is difficult. It measures the

logarithmic likelihood that runs of patterns that are close remain close on subsequent incremental comparisons, and assigns a non-negative number to a time series, with larger values corresponding to more complexity or irregularity in the data. ApEn has two user-specified parameters: a run length $m$ and a tolerance window $r$. It is important to consider ApEn($m$, $r$), or ApEn($m$, $r$, $N$), where $N$ is the number of points of the time series, as a family of parameters: comparisons between time series segments can only be made with the same values of $m$ and $r$ [45]. Formally, given $N$ data points from a time series $\{x(1), x(2),..., x(N)\}$, the ApEn is computed according to:

- From a sequence of $m$-vectors $X(1)$, $X(2)$,...,$X(N - m + 1)$ defined as $X(i) = [x(i), x(i + 1),...,x(i + m - 1)]$, $i = 1,2,...,N - m + 1$. These vectors represent $m$ consecutive $x$ values, commencing with the $i$th point.
- Calculate the distance between $X(i)$ and $X(j)$, $d[X(i),X(j)]$, as the maximum absolute difference between their respective scalar components,

$$d[X(i), X(j)] = \max_{k=0,1,\ldots,m-1}(|x(i+k) - x(j+k)|) \tag{12}$$

- For a given $X(i)$, count the number of $j$ ($j = 1,2,...,N - m + 1$) for $j \neq i$ such that $d = [X(i),X(j)] \leq r$, denoted as $N^m(i)$. Then, for $i = 1,2,...,N - m + 1$,

$$C_r^m(i) = \frac{N^m(i)}{N - m + 1} \tag{13}$$

The $C_r^m(i)$ values measure, within a tolerance $r$, the regularity (frequency) of patterns similar to a given segment of length $m$.

- Compute the natural logarithm of each $C_r^m(i)$, and compute the average of it over $i$,

$$\phi^m(r) = \frac{1}{N - m + 1}\sum_{l=1}^{N-m+1} \ln C_r^m(i) \tag{14}$$

where $\phi^m(r)$ represents the average frequency of all the $m$-point patterns in the sequence remain close to each other.

- Increase the dimension to $m + 1$. Repeat steps (1) to (4) and find $C_r^{m+1}$ and $\phi^{m+1}(r)$.
- Theoretically, the ApEn is defined as,

$$\mathrm{ApEn}(m, r) = \lim_{N\to\infty}\left[\phi^m(r) - \phi^{m+1}(r)\right]. \tag{15}$$

There are two ways to look at ApEn. From one point of view, it is a statistical metric (the average of the logarithm of a conditional probability), which makes it applicable to both deterministic and stochastic processes. From another point of view, it reflects the rate of new pattern generation and is thus related to the concept of entropy [22].

In practice, the number of data points $N$ is finite. We implemented this formula by defining the statistic [43]:

$$\mathrm{ApEn}(m, r) = \phi^m(r) - \phi^{m+1}(r). \tag{16}$$

In this study, we calculated the estimated ApEn of the interbeat interval series $\delta(n)$. ApEn was estimated with the established parameter values of $m = 1$ and $r = 0.25\ \sigma_\delta$, where $\sigma_\delta$ is the sample standard deviation of $\delta(n)$. Normalizing $r$ in this manner gives ApEn a translation and scale invariance, in that it remains unchanged under uniform process magnification, reduction, or constant shift to higher or lower values [45]. Several studies have demonstrated that these input parameters produce good statistical reproducibility for ApEn [27, 43, 48].

## 3 Duration/reliability analysis

### 3.1 Duration analysis

#### *3.1.1 Patient population*

We used the normal sinus rhythm R–R interval database posted on PhysioNet [40]. The database includes the beat times of 54 long-term ambulatory ECG recordings (21.4–24.2 h, 1,280 h total). The recordings were acquired from 30 men (ages 28.5–76 years) and 24 women (ages 58–73 years). The original recordings were sampled at a rate of 128 Hz. The beat and rhythm annotations were generated by an automatic algorithm and then manually reviewed and corrected.

#### *3.1.2 Record sampling*

We randomly selected 1,000 15-min records with replacement from the database. For each case, we randomly selected a subject and a 15 min record from the entire recording. We then discarded all records did not consist entirely of normal sinus rhythm beats. This reduced the number of usable records to 696.

#### *3.1.3 Figure of merit*

We calculated each of the 11 HRV metrics using segments of each 15 min ECG record. The segment durations ranged from 10 s to 10 min. We treated a

5 min segment in the center of each 15 min record (300–600 s) as a short-term ECG baseline record. We randomly selected segments of 5 min or less from the 5 min baseline. We selected segments of more than 5 min randomly from the 15 min record, but the selection was constrained to include all of the 5 min baseline. For each duration, we calculated the percent difference between the HRV metric calculated from the randomly selected segment and the 5 min baseline,

$$\epsilon(i,\tau) = 100 \times \frac{m(i,\tau) - m_{\mathrm{BL}}(i)}{\sigma_{\mathrm{BL}}} \tag{17}$$

where $\epsilon(i,\tau)$ is the percent difference, $i \in \{1, ..., 696\}$ is the record index, $\tau$ is the duration of the randomly selected segment, $m(i,\tau)$ is the metric calculated from the randomly selected segment, $m_{\mathrm{BL}}(i)$ is the metric calculated from the baseline, and $\sigma_{\mathrm{BL}}$ is the sample standard deviation of the metric calculated from the baselines of all 696 records.

## 3.2 Reliability analysis

### 3.2.1 Patient population

We used three data bases posted on PhysioNet [40]. The databases included the normal sinus rhythm RR interval database described previously, the European ST-T database, and the cardiac arrhythmia suppression trail (CAST) database. The European ST-T database consists of 90 annotated ambulatory recordings from 79 subjects. Each subject was diagnosed or suspected of having myocardial ischemia. The database includes patients with hypertension, ventricular dyskinesia. Some subjects were taking medication that may have affected their recordings. The original signals were sampled at 250 Hz.

The CAST database consisted of 762 pre-treatment recordings from patients who had an acute myocardial infarction within the preceding 2 years. Most of the signals were sampled at 128 Hz. All others were sampled at 125.4 Hz.

All three databases contained annotations that were generated by an automatic algorithm and then manually reviewed and corrected by one or more experts.

### 3.2.2 Record sampling

For each recording, we randomly selected a 15-min segment. If the segment did not consist entirely of normal sinus rhythm beats, we resampled the recording. Up to 500 attempts were made to find a 15-min segment with normal sinus rhythm beats. This resulted in fifty 15-min records out of 54 possible for the Normal Sinus Rhythm database, 41 out of 90 possible for the European ST-T database, and 480 out of 762 possible for the CAST database.

### 3.2.3 Intraclass correlation coefficients

Reliability analysis is a means of assessing what fraction of the variability in observed values is due to measurement error. We used intraclass correlation coefficients (ICC) as our measure of reliability [38, 53]. The ICC is defined as a ratio of explained variation to total variation. The calculation of these coefficients requires repeated measurements of the same entity, but allows for random errors in each measurement. There are several types of ICC that one may choose from depending on the statistical model of the study. We chose a one-way random effects model that can be estimated with an analysis of variance (ANOVA) model II with random factor levels. This treats each observation as a realization of the following random process

$$y_{ij} = \mu_i + \varepsilon_{ij} \tag{18}$$

where $i$ is the record index, $j$ is the measurement index, $\mu_i$ are independent and normally distributed $N(\mu_.,\sigma_\mu^2)$, $\varepsilon_{ij}$ are independent and normally distributed $N(0,\sigma^2)$, and $\mu_i$ and $\varepsilon_{ij}$ are mutually independent.

We treated the metrics calculated from the first (0–300 s) and last (600–900 s) 5 min segments of each 15 min record as a repeated measurement of the same entity. The ICC was calculated as,

$$\hat{\rho} = \frac{\mathrm{MSTR} - \mathrm{MSE}}{\mathrm{MSTR} + (M-1)\mathrm{MSE}} = \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + \hat{\sigma}^2} \tag{19}$$

where $M = 2$ is the number of measurements per entity, MSTR is the treatment mean square, and MSE is the error mean square. This is an estimate of the true ICC defined as

$$\rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}. \tag{20}$$

The estimate is biased, but consistent and tolerant of modest departures from normality [53]. Values close to one indicate that the variation between entities is significantly greater than measurement error and values close to zero indicate that the measurement error is too large to accurately discern differences between entities.

## 4 Results and discussion

Our results indicate that all the HRV metrics included in this study are sensitive to changes in duration of the ECG segment (Fig. 1). In these plots we show the median and four ranges of the percent difference in (17) for ECG segment durations ranging from 10 s to 10 min. From these results we also conclude that most of the HRV metrics are biased estimates. This confirms the task force's statement that metrics calculated from segments of different durations are not comparable [57].

The plots in Fig. 1 indicate that all the HRV metrics are sensitive to changes in the duration of the ECG segment. Most of the total HRV is caused by drift and nonstationary effects that cannot be estimated accurately from short segments. HF, SDSD, and RMSSD had the best overall performance because they both effectively highpass filter the NN interval series and are thereby less sensitive to these effects than the other metrics. The bias and variance of HF, SDSD, and RMSSD were small and comparable to the mean heart rate (MHR).

The results of the reliability study (Tables 1, 2, 3) demonstrate that the MHR was much more reproducible than any of the other HRV metrics. RMSSD, SDSD, and HF were significantly more reliable than the other metrics and achieved an ICC comparable to MHR. ApEn was the most unreliable metric of the HRV measures tested in this study. We also note that pairs of 5 min segments spaced 5 min apart are not actually repeated measurements of the same entity since the NN interval series is known to be a nonstationary process. Here we assumed that the process is locally stationary and that the metrics calculated from each pair of segments can be treated as independent samples drawn from the same distribution.

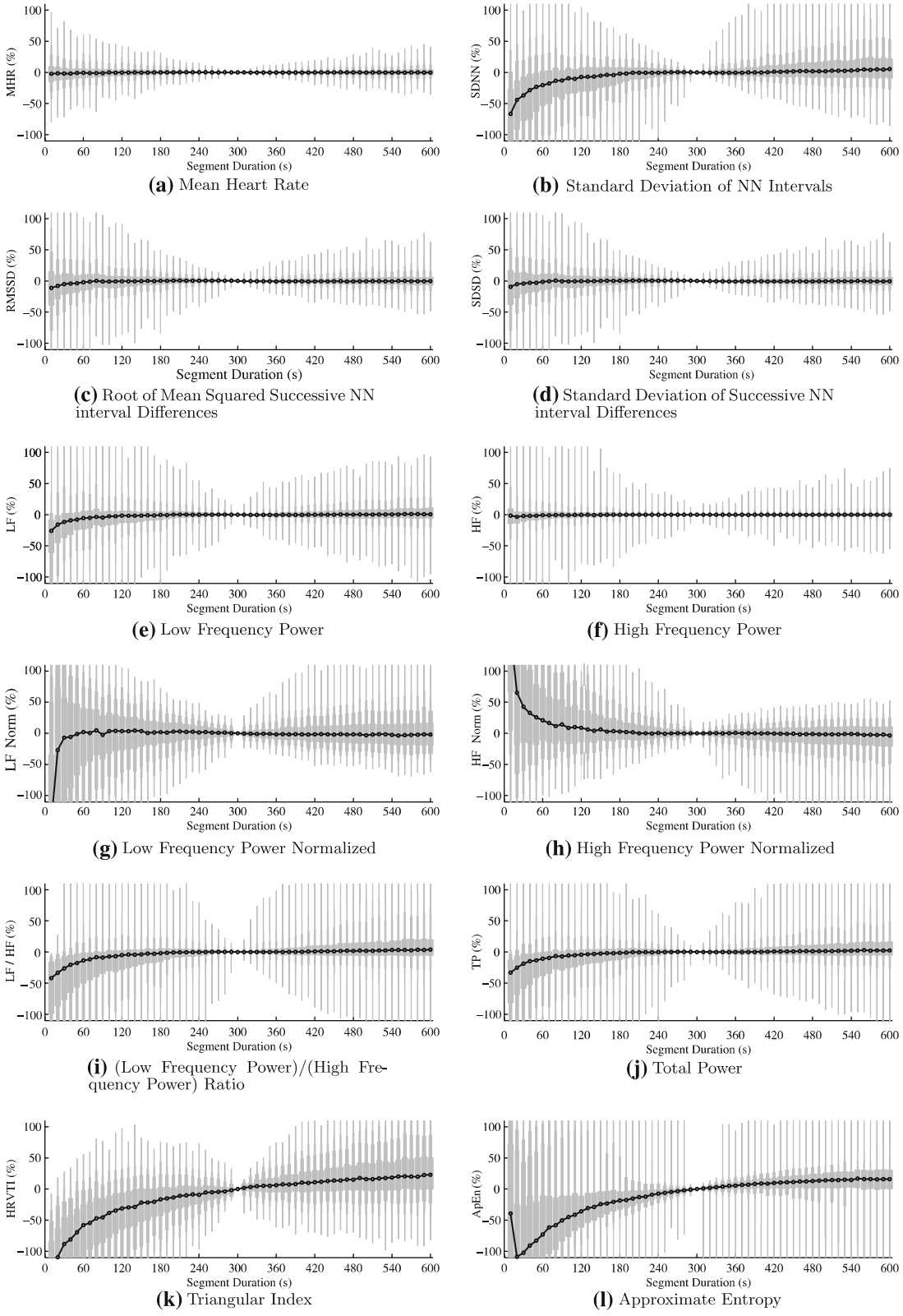
It is important to recognize that reliability depends on the variability between subjects, that is, the second column of Tables 1, 2, and 3 varies across databases. This may explain the discrepancy in the ICCs across the databases, and the discrepancy of the results of reliability studies on HRV metrics reported in the literature. However, the first column should be reasonably consistent across databases. Additionally, the precise measurement of the NN intervals also affect the reliability of the HRV metrics [4, 10, 11, 12, 15, 16].

Our results complement those of several other research groups that have studied the reproducibility of HRV metrics in a variety of situations. In general, the objective of previous works has been to assess the reproducibility of a subset of the HRV metrics used to study a specific patient population or condition of interest. These studies include research on the short- and long-term reproducibility of autonomic measures in supine and standing positions [31], reproducibility of HRV responses to graded lower body negative pressure [20, 34], reliability of short-term HRV measures during exercise [3, 28, 30, 62], reproducibility of measures of cardiovascular autonomic nervous function in middle age and elderly subjects [23], reproducibility of frequency domain HRV metrics before and after a standardized meal [18], HRV analysis reproducibility in the chronic phase of myocardial infarction [5], reproducibility of HRV from short-term recording during manoeuvres in normal subjects [9], stability of short recordings in time [54] and reproducibility of HRV metrics obtained from short-term sampling records [21, 37]. In addition to the reproducibility of 24 h and 5 min records, assessing the reproducibility of HRV metrics calculated from very short records has significant practical importance [25], since it is not always possible to obtain 5 min records due to instrumentation constraints or study design [17]. For example, many 12-lead ECG instruments acquire record of only 10 s in duration.

The main limitation of this work is that we analyzed data from ambulatory recordings, possibly not collected under stationary conditions as recommended by the task force. This may partly explain the lower reliability of the normal sinus rhythm database metrics. These subjects may also have been more active than the other groups.

## 5 Conclusion

We estimated the reliability of 11 metrics calculated from 5 min records and compared the accuracy of the HRV metrics calculated from ECG records spanning 10 s to 10 min to that calculated from a 5 min record and proposed a new methodology and visualization technique which can be used to assess the segment length required for a given degree of accuracy. Our results indicate that all the HRV metrics are sensitive to changes in the duration of the ECG segment. Most of the total HRV is caused by drift and nonstationary effects that cannot be estimated accurately from short segments. HF, SDSD, and RMSSD had the best overall performance because they both effectively highpass filter the NN interval series and are thereby less sensitive to these effects than the other metrics. The results of the reliability study demonstrate that the MHR was much more reproducible than any of the other HRV metrics.




**(a)** Mean Heart Rate

**(b)** Standard Deviation of NN Intervals

**(c)** Root of Mean Squared Successive NN interval Differences

**(d)** Standard Deviation of Successive NN interval Differences

**(e)** Low Frequency Power

**(f)** High Frequency Power

**(g)** Low Frequency Power Normalized

**(h)** High Frequency Power Normalized

**(i)** (Low Frequency Power)/(High Frequency Power) Ratio

**(j)** Total Power

**(k)** Triangular Index

**(l)** Approximate Entropy

**Fig. 1** Plots of the difference between eleven different heart rate metrics calculated from a range of segment durations and a 5 min baseline segment. This figure shows the mean heart rate and the other eleven plots show HRV metrics. Each plot shows the median difference and three ranges calculated from 696 15-min records. The differences are shown as a percentage of the standard deviation of metric calculated from the 5 min baseline of all 696 records. The *gray vertical bars* show four different ranges of each metric: 25–75, 10–90, 5–95, 1–99%

**Table 1** Summary of a reliability analysis of the cardiac arrhythmia suppression trail (CAST) database

| Metric | $\hat{\sigma}$ | $\hat{\sigma}_\mu$ | 2.5% | $\hat{\rho}$ (%) | 97.5% |
|---|---|---|---|---|---|
| MHR (Hz) | 0.001 | 0.054 | 97.6 | 98.0 | 98.3 |
| SDNN (ms) | 0.123 | 0.281 | 64.5 | 69.5 | 73.8 |
| HRVTI | 3.419 | 9.250 | 68.5 | 73.0 | 76.9 |
| RMSSD (ms) | 0.020 | 0.188 | 88.8 | 90.6 | 92.1 |
| SDSD (ms) | 0.020 | 0.189 | 88.8 | 90.6 | 92.1 |
| LF ($ms^2$) | 0.088 | 0.236 | 68.3 | 72.8 | 76.7 |
| HF ($ms^2$) | 0.005 | 0.047 | 87.8 | 89.7 | 91.3 |
| LF Norm | 0.006 | 0.023 | 76.8 | 80.2 | 83.2 |
| HF Norm | 0.004 | 0.014 | 72.3 | 76.3 | 79.8 |
| LF/HF | 0.000 | 0.000 | 30.8 | 38.7 | 46.0 |
| TP ($ms^2$) | 1.758 | 1.994 | 46.4 | 53.1 | 59.3 |
| ApEn | 0.052 | 0.071 | 51.5 | 57.8 | 63.4 |

**Table 2** Summary of a reliability analysis of the European ST-T Database

| Metric | $\hat{\sigma}$ | $\hat{\sigma}_\mu$ | 2.5% | $\hat{\rho}$ (%) | 97.5% |
|---|---|---|---|---|---|
| MHR (Hz) | 0.001 | 0.061 | 98.1 | 99.0 | 99.4 |
| SDNN (ms) | 0.087 | 0.624 | 78.4 | 87.8 | 93.3 |
| HRVTI | 2.403 | 11.418 | 69.8 | 82.6 | 90.3 |
| RMSSD (ms) | 0.021 | 0.210 | 83.5 | 90.8 | 95.0 |
| SDSD (ms) | 0.021 | 0.211 | 83.5 | 90.8 | 95.0 |
| LF ($ms^2$) | 0.342 | 1.150 | 61.1 | 77.1 | 87.0 |
| HF ($ms^2$) | 0.019 | 0.069 | 63.1 | 78.3 | 87.8 |
| LF Norm | 0.006 | 0.024 | 64.7 | 79.4 | 88.4 |
| HF Norm | 0.003 | 0.016 | 71.5 | 83.7 | 90.9 |
| LF/HF | 0.000 | 0.000 | 74.2 | 85.3 | 91.9 |
| TP ($ms^2$) | 2.680 | 13.066 | 70.4 | 83.0 | 90.5 |
| ApEn | 0.046 | 0.082 | 41.8 | 63.9 | 78.9 |

**Table 3** Summary of a reliability analysis of the normal sinus rhythm RR interval database

| Metric | $\hat{\sigma}$ | $\hat{\sigma}_\mu$ | 2.5% | $\hat{\rho}$ (%) | 97.5% |
|---|---|---|---|---|---|
| MHR (Hz) | 0.002 | 0.031 | 87.3 | 92.5 | 95.7 |
| SDNN (ms) | 0.204 | 0.199 | 25.4 | 49.4 | 67.7 |
| HRVTI | 4.076 | 4.915 | 32.0 | 54.7 | 71.4 |
| RMSSD (ms) | 0.056 | 0.135 | 53.5 | 70.6 | 82.1 |
| SDSD (ms) | 0.056 | 0.135 | 53.5 | 70.6 | 82.1 |
| LF ($ms^2$) | 0.151 | 0.231 | 39.5 | 60.4 | 75.4 |
| HF ($ms^2$) | 0.013 | 0.067 | 72.5 | 83.4 | 90.2 |
| LF Norm | 0.007 | 0.024 | 63.7 | 77.6 | 86.6 |
| HF Norm | 0.005 | 0.017 | 62.2 | 76.6 | 86.0 |
| LF/HF | 0.000 | 0.000 | 0.0 | 14.0 | 39.9 |
| TP ($ms^2$) | 2.879 | 1.833 | 12.9 | 38.9 | 60.0 |
| ApEn | 0.099 | 0.039 | 0.6 | 28.0 | 51.5 |

**Acknowledgments** We are grateful to Phyllis Stein and Rochelle Goldsmith for contributing their data to PhysioBank, and to the anonymous referees for their thorough revision of the original manuscript and helpful suggestions.

## References

1. Akselrod S, Gordon D, Ubel FA, Shannon DC, Barger CA, Cohen RJ (1981) Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardovascular control. Science 213(4504):220–222
2. Amara C, Wolfe L (1998) Reliability of noninvasive methods to measure cardiac autonomic function. Can J Appl Physiol 23(4):396–408
3. Bernardi L, Valle F, Coco M, Calciati A, Sleight P (1996) Physical activity influences heart rate variability and very-low frequency components in Holter electrocardiogram. Cardiovasc Res 32(2):234–227
4. Bragge T, Tarvainen MP, Ranta-Aho PO, Karjalainen PA (2005) High-resolution QRS fiducial point corrections in sparsely sampled ECG recordings. Physiol Meas 26(5):743–751 [Online]. Available: http://dx.doi.org/10.1088/0967-3334/26/5/013
5. Brembilla-Perrot B, Houriez P, Jacquemin L, Houplon P, Claudon O, Danchin N (2004) Reproducibility of heart rate variability in chronic phase myocardial infaction. Arch Mal Coeur Vaiss 2(2):245–252
6. Breuer H, Skyschally A, Wehr M, Schulz R, Heusch G (1992) Poor reproducibility of heart rate variability indices. Z Kardiol 81(9):475–481
7. Buchman TG, Stein PK, Goldstein B (2002) Heart rate variability in critical illness and critical care. Curr Opin Critic Care 8(4):311–315
8. Burger A, Charlamb M, Weinrauch L, DElia J (1997) Short- and long-term reproducibility of heart rate variability in patients with long-standing type I diabetes mellitus. Am J Cardiol 80(9):1198–1202
9. Carrasco S, Gonzßlez R, Gaitßn MJ, Yez O (2003) Reproducibility of heart rate variability from short-term recordings during five manoeuvres in normal subjects. J Med Eng Technol 27(6):241–248 [Online]. Available: http://dx.doi.org/10.1080/0309190031000111380
10. Christov II (2004) Real time electrocardiogram QRS detection using combined adaptive threshold. Biomed Eng Online 3(1):28 [Online]. Available: http://dx.doi.org/10.1186/1475-925X-3-28
11. Christov II, Daskalov IK (1999) Filtering of electromyogram artifacts from the electrocardiogram. Med Eng Phys 21(10):731–736
51 Christov II, Dotsinsky IA, Daskalov IK (1992) High-pass filtering of ECG signals using QRS elimination. Med Biol Eng Comput 30(2):253–256
13. Cloarce-Blanchard L, Funck-Brentano C, Lipski M, Jaillon P, Macquin-Mavier I (1997) Repeatability of spectral components of short-term blood pressure and heart rate variability during acute sympathetic activation in healthy young male subjects. Clin Sci (Lond) 93(1):21–28
14. D'Addio G, Acanfora D, Pinna G, Maestri R, Furgi G, Piconne C, Rengo F (1998) Reproducibility of short- and long-term Poincare plot parameters compared with frequency-domain hrv indexes in congestive heart failure. Comput Cardiol pp 381–384
15. Daskalov I, Christov I (1997) Improvement of resolution in measurement of electrocardiogram RR intervals by interpolation. Med Eng Phys 19(4):375–379
16. Daskalov IK, Christov IW (1999) Electrocardiogram signal preprocessing for automatic detection of QRS boundaries. Med Eng Phys 21(1):37–44

17. Dekker JM, Schouten EG, Klootwijk P, Pool J, Swenne CA, Kromhout D (1997) Heart rate variability from short electrocardiographic recordings predicts mortality from all causes in middle-aged and elderly men. Am J Epidemiol 145(10):899–908
18. Dionne I, White M, Tremblay M (2002) The reproducibility of power spectrum analysis of heart rate variability before and after a standardized meal. Physiol Behav 75(3):267–270
19. Duanping L, Barnes R, Chambless L, Heiss G (1996) A computer algorithm to impute interrupted heart rate data for the spectral analysis of heart rate variability—the aric study. Comput Biomed Res 29(2):140–151
20. Franke W, Buchanan D, Lee K (2001) Reproducibility of the autonomic and cardiovascular responses to maximam lower body negative pressure. Med Sci Sports Exerc 33(5):S209
21. Freed L, Stein K, Gordong M, Urban M, Kligfield P (1994) Reproducibility of power spectral measures of heart rate variability obtained from short-term sampling periods. Am J Cardiol 74(9):972–973
22. Fusheng Y, Bo H, Qingyu T (2001) Approximate entropy and its application in biosignal analysis. In: Akay M (ed) Nonlinear biomedical signal processing, vol II: dynamic analysis and modeling, ser. Biomedical engineering. IEEE Press, New York
23. Gerritsen J, TenVoorde B, Decker J, Kingma R, Kostense P, Bouter L, Heethaar R (2003) Measures of cardiovascular autonomic nernous function: agreement, reproducibility, and reference values in middle age and elderly subjects. Diabetologica 46(3):330–338
24. Ghuran A, Malik M (1999) Heart rate variability—state of the art. Cardiac Electrophysiol Rev 3:283–285
25. Hamilton R, Mckenchnie P, Macfarlane P (2004) Can cardiac vagal tone be estimated from the 10-second ECG? Int J Cardiol 95(1):109–115
26. Hedman AE, Hartikainen JEK (1999) Has non-linear analysis of heart rate variability any practical value? Cardiac Electrophysiol Rev 3:286–289
27. Kaplan D, Furman M, Pincus S, Ryan S, Lipsitz L, Goldberger A (1991) Aging and the complexity of cardiovascular dynamics. Biophys J 59:945–949
28. Kerrigan D, Armstrong WJ, Levine S, Ehrman J (2001) Reproducibility of heart rate variability during exercise. Med Sci Sports Exerc 59(44):S202
29. Klingenheben T, Zabel M, Just H, Hohnloser S (1993) Reproducibility of heart rate variability measures as determined from repeated holter monitorings. Z Kardiol 82(5):302–308
30. Kluess H, Wood R, Stone D, Weslch M (2001) Reliability of heart rate variability during dynamic handgrip exercise. Med Sci Sports Exerc 33(5):S203
31. Kowalewski M, Urban M (2004) Short- and long-term reproducibility of autonomic measures in supine and standing positions. Clin Sci (Lond) 106(1):61–66
32. Lake D, Richman J, Griffin M, Moorman J (2002) Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol 283:R789–R797
33. Lawrence G, Home P, Murray A (1992) Repeatability of measurements and sources of variability in tests of cardiovascular autonomic function. Br Heart J 68(2):205–211
34. Lee K, Buchanan D, Flatau A, Franke W (2004) Reproducibility of the heart rate variability responses to graded lower negative pressure. Eur J Appl Physiol 92(1-2):106–113
35. Liao D, Barnes R, Chambless L, Heiss G (1996) A computer algorithm to impute interrupted heart rate data for the spectral analysis of heart rate variability: the aric study. Comput Biomed Res 29(2):140–151
36. Malpas SC (2002) Neural influences on cardiovascular variability: possibilities and pitfalls. Am J Physiol Heart Circ Physiol 282:H6–H20
37. Marks B, Lightfoot J (1999) Reproducibility of resting heart rate variability with short sampling periods. Can J Appl Physiol 24(24):337–348
38. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods 1(1):30-46; corrections in no. 4, p 390
39. Mehta S, Super D, Salvator A, Fradley L, Connuck D, Kaufman E (2002) Heart rate variability by triangular index in infants exposed prenatally to cocaine. Ann Noninvasive Electrocardiol 7(4):374–378
40. Moody GB, Mark RG, Goldberger AL (2001) PhysioNet: a web-based resource for the study of physiologic signals. IEEE Eng Med Biol Mag 20(3):70–75
41. Myrtek M (1990) Covariation and reliability of ECG parameters during 24-hour monitoring. Int J Psychophysiol 10(2):117–124
42. Palazzolo J, Estafanous F, Murray P (1998) Entropy measures of heart rate variation in conscious dogs. Am J Heart Circ Physiol 274(4):H1099–H1105
43. Pincus S (1991) Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA 88:2297–2301
44. Pincus S (1992) Approximating markov chains. Proc Natl Acad Sci USA 89:4432–4436
45. Pincus S (2001) Assessing serial irregularity and its implications for health. Ann N Y Acad Sci 954:245–267
46. Pincus S, Goldberger A (1994) Physiological time series analysis: what does regularity quantify? Am J Physiol Heart Circ Physiol 266:H1643–H1656
47. Pincus S, Kalman R (1997) Not all (possibly) ''random'' sequences are created equal. Proc Natl Acad Sci USA 94(8):3513–3518
48. Pincus S, Keefe D (1992) Quantification of hormone pulsatility via an approximate entropy algorithm. Am J Physiol Endocrinol Metab 262:E741–E754
49. Pincus S, Singer B (1996) Randomness and degrees of irregularity. Proc Natl Acad Sci USA 93:2083–2088
50. Pitzalis M, Mastropasqua F, Massari F, Forleo C, DiMaggio M, Passantino A, Colombo R, DiBiase M, Rizzon P (1996) Short- and long-term reproducibility of time and frequency domain heart rate variability measurements in normal subjects. Cardiovasc Res 32(2):226–233
51. Ponikowski P, Piepoli M, Amandi A, Chua T, Harrington D, Volterrani M, Colombo R, Mazzuero G, Giordano A, Coats A (1996) Reproducibility of heart rate variability measures in patients with chronic failure. Clin Sci (Lond) 91(4):391–398
52. Saul JP (1990) Beat-to-beat variations of heart rate reflect modulation of cardiac autonomic outflow. News Physiol Sci 5:32–37
53. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420–428
54. Sinnreich R, Kark JD, Friedlander Y, Sapoznikov D, Luria MH (1998) Five minute recordings of heart rate variability for population studies: repeatability and age–sex characteristics. Heart 80(2):156–162
55. Stein P, Rich M, Rottman J, Kleiger R (1995) Stability of index of heart rate variability in patients with congestive heart failure. Am Heart J 129(5):975–981
56. Taverner D, Nunan T, Tonkin A (1996) Reproducibility of of conventional and power spectral measurements of cardiovascular sympathetic activation in normal subjects. Clin Exp Pharmacol Physiol 23(9):804–806

57. TF of the European Society of Cardiology, the North American Society of Pacing, and Electrophysiology (1996) Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93:1043–1065
58. Toyry J, Mantysaari M, Hartikainen J, Lansimies E (1995) Day-to-day variability of cardiac autonomic regulation parameters in normal subjects. Clin Physiol (Oxford) 15(1):39–46
59. vandeBorne P, Montano N, Zimmerman B, Pagani M, Somers V (1997) Relationship between repeated measures of hemodynamics, muscle sympathetic nerve activity, and their oscillations. Circulation 96(12):4326–4332
60. Vanhoogenhuyze D, Weinstein N, Martin G, Weiss J, Schaad J, Sahyouni X, Fintel D, Remme W, Singer D (1991) Reproducibility and relation to mean heart rate of heart rate variability in normal subjects. Am J Cardiol 68(17):1668–1676
61. Vardas P, Kochiadakis G, Orfanakis A, Kalaitzakis M, Manios E (1994) Intraindividual reproducibility of heart-rate-variability before and during postural tilt in patients with syncope of unknown origin. Pacing Clin Electrophysiol 17(11):2207–2210
62. Winsley R, Armstrong N, Bywater K, Fawkner S (2003) Reliability of heart rate variability measures at rest and during light exercise in children. Br J Sports Med 37(6):550–552