Int J Psychophysiol. Author manuscript; available in PMC 2017 July 07.

Published in final edited form as:
Int J Psychophysiol. 2012 October ; 86(1): 88–97. doi:10.1016/j.ijpsycho.2012.07.004.

Measurement fidelity of heart rate variability signal processing: The devil is in the details

Denise C. Jarrin[a], **Jennifer J. McGrath**[a,*], **Sabrina Giovanniello**[a], **Paul Poirier**[b], and **Marie Lambert**[c,1]

[a]Pediatric Public Health Psychology Laboratory, Department of Psychology, Concordia University, 7141 Sherbrooke St. W, Montréal, QC, Canada, H4B 1R6

[b]Faculty of Pharmacy, Laval University, Quebec Heart and Lung Institute, 2725 Chemin Sainte-Foy, Québec QC, Canada, G1V 4G5

[c]Centre de recherche du CHU Sainte-Justine, 3175 Ch. de la Cote-Sainte-Catherine, Montreal, QC, Canada, H3T 1C5

Abstract

Heart rate variability (HRV) is a particularly valuable quantitative marker of the flexibility and balance of the autonomic nervous system. Significant advances in software programs to automatically derive HRV have led to its extensive use in psychophysiological research. However, there is a lack of systematic comparisons across software programs used to derive HRV indices. Further, researchers report meager details on important signal processing decisions making synthesis across studies challenging. The aim of the present study was to evaluate the measurement fidelity of time- and frequency-domain HRV indices derived from three predominant signal processing software programs commonly used in clinical and research settings. Triplicate ECG recordings were derived from 20 participants using identical data acquisition hardware. Among the time-domain indices, there was strong to excellent correspondence (ICC_{avg} =0.93) for SDNN, SDANN, SDNNi, rMSSD, and pNN50. The frequency-domain indices yielded excellent correspondence (ICC_{avg} =0.91) for LF, HF, and LF/HF ratio, except for VLF which exhibited poor correspondence (ICC_{avg} =0.19). Stringent user-decisions and technical specifications for nuanced HRV processing details are essential to ensure measurement fidelity across signal processing software programs.

Keywords

Heart rate variability; Measurement fidelity; Signal processing; Psychometrics; Reliability

*Corresponding author at: Department of Psychology, Concordia University, 7141 Sherbrooke St. West, Montréal, QC Canada, H4B 1R6.

[1]Posthumous. Special acknowledgement goes to Dr. Marie Lambert (deceased February 20, 2012), for without her vision and dedication, Team PRODIGY and the QUALITY cohort would not exist.

1. Introduction

Heart rate variability (HRV) is an indicator of the total amount of oscillations of heart periods between consecutive QRS complexes of normal sinus depolarizations (RR intervals). Reduced HRV, suggested to reflect hyperactive sympathetic and/or hypoactive parasympathetic nervous system activity, has been implicated in the pathophysiology of a number of health outcomes including cardiac conditions such as myocardial infarction, coronary heart disease, and hypertension (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996; Liao et al., 2002), and non-cardiac conditions such as obesity, diabetes (Masi et al., 2007), insulin resistance (Lindmark et al., 2003), metabolic syndrome (Hemingway et al., 2005), dyspepsia (Lorena et al., 2002), irritable bowel syndrome, anorexia nervosa (Mazurak et al., 2011), epilepsy (Ferri et al., 2002), anxiety (Friedman, 2007; Friedman and Thayer, 1998), and major depressive disorder (Nugent et al., 2011), as well as mortality (Camm et al., 2001; Gerritsen et al., 2001; Thayer and Lane, 2007). Significant developments in statistical, spectral, and geometric signal processing to automatically derive HRV parameters have led to their increased use in multidisciplinary settings. As such, many signal processing software programs have been created to analyze HRV data. These programs offer rapid, automatic analysis of output based on sophisticated signal processing techniques and algorithms that identify and measure various electrocardiograph (ECG)-derived variables from each cardiac cycle. While there are previous recommendations from the Task Force (1996) for comparative data across studies, there is a lack of systematic comparisons across computer software programs used to derive common time-and frequency-domain HRV indices.

1.1. Heart rate variability

Traditionally, the autonomic nervous system (ANS) has been thought to be reciprocally balanced (i.e., as one branch of the ANS increases activity the other branch decreases activity); however, evidence suggests that parasympathetic and sympathetic outflows are distributed multidimensionally (Berntson et al., 1991). As such, HRV and each of its components are particularly valuable quantitative markers that provide information on the flexibility and balance of the branches of the ANS based on heart period series (Berntson et al., 1997; Task Force, 1996).

Although heart period series can be construed from several physiological signals including photoplethysmography (Lu et al., 2009), continuous blood pressure recordings (Parati et al., 1995), doppler ultrasound techniques (Jezewski et al., 2008), and microwave reflectometry (Mase et al., 2010), it is most typically derived from continuously recorded ECG signals. Many of these alternative physiological signals yield only approximate indicators of heart period series (Berntson et al., 1997). For example, ambiguous waveform morphology from distal photoplethysmographic records or continuous blood pressure recordings contribute to difficulty identifying accurate reference points (Berntson, et al, 1997). ECG recordings are preferred and considered a simple, noninvasive technique with clear waveform morphology, as the instantaneous ventricular depolarization yields the highest signal-to-noise ratio rendering a clearly delineated R-wave (Berntson et al., 1997; Lu et al., 2009; Task Force,

1996). Further, ECG recordings provide proximal and reliable information on heart period series to quantify HRV and ultimately, evaluate autonomic function and balance.

Following data acquisition and audio-to-digital (A/D) conversion of raw ECG signals, HRV analysis is comprised of two major phases: signal preprocessing and automated analyses to derive HRV parameters (Berntson et al., 1997; Kligfield et al., 2007). Signal preprocessing incorporates accurately identifying QRS complexes and removing artifacts, while still preserving the integrity of the respiratory sinus rhythm. Artifacts may be attributable to movement (i.e., muscle activity), external electromagnetic signals (e.g., 50/60 Hz power lines), or technical problems (e.g., poorly fastened electrodes; Berntson et al., 1997; Berntson and Stowell, 1998). Failure to identify artifacts can lead to missing or additional QRS complex detections and minor contamination can increase error in HRV results by up to 30% (Berntson et al., 1997; Berntson and Stowell, 1998; Xia et al., 1993). Signal preprocessing is influenced by both technical specifications (e.g., sampling rate, digital filters; Bailey et al., 1990; Mortara, 1977; van Bemmel et al., 1990) and algorithms used for ECG pattern recognition and interpolation (e.g., feature extraction, beat selection; Bailey et al., 1974; Bonner and Schwetman, 1968; Pipberger et al., 1962). Detector algorithms can be based on heuristic derivative equations that identify discrete measurements or adaptive thresholds, for example, the increasing edge of the R-peak (Bonner et al., 1972; Pryor et al., 1969). Alternatively, they can be based on complex statistical algorithms that use linear or nonlinear filters, different transformations, or discriminant function analysis (Köhler et al., 2003; Pan and Tompkins, 1985; Romhilt and Estes, 1968). Interpolation algorithms, to replace missing or abnormal heart period series, include proximal, piecewise cubic Hermite, non-linear predictive interpolation, linear, and cubic spline interpolations (Kim et al., 2009; Lippman et al., 1994; Malik and Camm, 1995).

Automated analyses predominantly use linear analyses such as time- and frequency-domain methods to quantify HRV indices (Task Force, 1996). Other nonlinear analyses including fractal (e.g., detrended fluctuation analysis, power-law correlation; Pincus, 1995; Richman and Moorman, 2000), symbolic dynamics (Porta et al., 2001; Voss et al., 2009), and complexity/entropy measures (e.g., approximate entropy, sample entropy, Shannon entropy, corrected conditional entropy, multiscale entropy), also exist (Montano et al., 2012; Porta et al., 2001; Voss et al., 2009). Although nonlinear analyses provide quantitative information on the regularity and complexity of autonomic cardiovascular control, linear analyses are most commonly reported in the literature. Time-domain approaches are based on statistical calculations derived from the direct measurement of RR intervals (e.g., SDNN, SDANN, SDNNi) or from the differences between successive RR intervals (e.g., rMSSD, pNN50; Task Force, 1996). Methodological study designs partly guide how data are partitioned for cleaning and aggregating. Partitioning data into meaningful conditions (e.g., baseline vs. task), categories (e.g., day vs. night), or smaller segments due to signal quality (e.g., 2 h segment vs. ten 20 min segments) interrupts the contiguous nature of the ECG signal. Further, data reduction decisions on the duration of analytical epochs (e.g., 1 vs. 5 min) to compute aggregated HRV indices across the epochs may yield different values. These decisions have important implications and must be carefully considered, especially for time-domain variables.

Frequency-domain variables are based on spectral analysis of RR intervals (Lahiri et al., 2008). Power spectral density decomposes RR intervals into their fundamental frequency components and provides information on the distribution of power as a function of frequency. Spectral analyses can include parametric (autoregressive; Yule–Walker, Burg) or nonpara-metric methods (Fast Fourier Transform, FFT; Kim et al., 2009). FFT is most commonly used to calculate the maximum variability in heart period series, based on ranges of frequency-specific oscillations of the RR intervals that reflect different branches of the cardiac system (Lahiri et al., 2008; Spiers et al., 1993).

Low frequency (LF) ranges from 0.04 to 0.15 Hz and reflects the aggregate influences of both sympathetic and parasympathetic branches of the ANS (Akselrod et al., 1981; Berntson et al., 1997); although, some researchers have suggested LF to be mainly of sympathetic origin (Malliani et al., 1991). High frequency (HF) ranges from 0.15 to 0.40 Hz and represents parasympathetic activity (Berntson et al., 1997; Pomeranz et al., 1985). Less studied frequencies include very low frequency (0.0033 to 0.04 Hz) and ultra low frequency (<0.003 Hz) ranges; these are thought to be influenced by the renin–angiotensin system as well as thermoregulatory processes and circadian rhythms (Kitney, 1980; Taylor et al., 1998). Importantly, default frequency bandwidths may differ across software programs leading to misinterpretation of the calculated HRV indices. For example, if HF was set incorrectly to 0.12 to 0.40 Hz, results would actually include LF as well, and therefore, not solely represent the parasympathetic nervous system.

Another important decision for spectral analyses includes windowing. Spectral windowing involves the application of a window function, of a specified width, to shape the time portion of ECG data by overlapping waveform endpoints in a smooth, continuous way without sharp transitions to minimize edge effects that result in spectral leakage for better overall spectral resolution. Hamming or Hanning windows are commonly used due to their high quality frequency resolution and reduced spectral leakage (Bloomfield, 1976; Harris, 1978).

In the extant literature, research studies that use HRV report meager details on the methodological decisions related to signal preprocessing specifications, algorithms, and interpolation methods used. Time- and frequency-domain HRV indices are vulnerable to artifacts, missing data, temporal factors, and trends in RR intervals (Kim et al., 2009, 2007; Spiers, et al., 1993; Task Force, 1996), and are thus highly influenced by decisions for data reduction, artifact detection and removal, and technical specifications (e.g., digital filtering, sampling frequency, detector or interpolation algorithms, windowing; Kim et al., 2009; Task Force, 1996; Welch, 1967).

Despite its extensive use, the comparability between standard computer HRV software programs has not been systematically evaluated. There is scant evidence of comprehensive comparisons to assess the fidelity of signal processing across multiple software programs. Of the only study to compare HRV signal processing programs, Jung et al. (1996) found time- and frequency-domain variables were not comparable across four programs in widespread use at the time almost two decades ago. Jung attributed the large variability across programs to different technical specifications, including beat selection methods (e.g.,

best complex, time-coherent averaging, extraction), sampling frequency, interpolation, and algorithms.

1.2. Present study

A significant challenge exists for researchers who want to compare or synthesize HRV results across studies. In a systematic review on short-term HRV measures, Nunan et al. (2010) found considerably large variations across studies, especially for frequency-domain variables. These discrepancies were attributed to differences in study design and methodology, as well as failure from authors to provide pertinent information on signal processing and data cleaning procedures. The importance of standardization across studies was reinforced by the Task Force (1996) guidelines in hopes to facilitate the exchange of knowledge, allow for comparative results across studies, and avoid conflicting data due to different technical and methodological approaches. Following these recommendations for standardization and interpretation of HRV measures, the purpose of the present study was to evaluate the measurement fidelity of HRV indices derived from three predominant signal processing software programs most commonly used in clinical and research settings among cardiologists, psychophysiologists, and other researchers across diverse disciplines (MARS, MindWare, Kubios). Using triplicate ECG data derived from identical data acquisition hardware, the comparability of HRV indices for time-domain (i.e., SDNN, SDANN, SDNNi, rMSSD, pNN50) and frequency-domain variables (LF, HF, LF/HF ratio) was tested.

2. Material and methods

2.1. Measures

2.1.1. ECG data acquisition—Twenty Holter tapes with raw ECG data were randomly chosen from an ongoing study of healthy youth participants between the ages of 8 and 11 (M_{age} =9.93 years, SD =1.02; 55% male). The complete research protocol is described elsewhere (Lambert et al., 2011). All ECG recordings were reviewed by a board-certified cardiologist; no cardiovascular pathology was identified (i.e., bradycardia, fibrillation, premature contraction). During the standardized protocol conducted in a hospital setting, continuous raw ECG data were acquired using the 8500 Marquette MARS Holter monitor (GE Marquette Medical Systems, Milwaukee, Wisconsin, USA), digitized (128 Hz), and recorded on a frequency modulated cassette recorder. The Holter monitor incorporated a quartz-derived, binary time channel that was automatically zeroed at the start of the recording. ECG acquisition began in the morning between 8 and 9 am and lasted approximately 2.5 h.

ECG data was derived from a modified Lead II configuration using disposable, pre-gelled snap silver chloride electrodes. Electrode resistance was minimized (<10 kO) by precleaning the skin with rubbing alcohol swabs. The active electrode (and its derivative/dZ) was placed on the right clavicle next to the sternum over the first rib between the two collarbones. The second electrode was placed on the left mid-clavicular line at the apex of the heart over the ninth rib. The ground electrode was placed near the lowest possible right rib cage on the abdomen. Additional dZ electrodes were placed over the right fourth intercostal space at the sternal edge, the fifth intercostal space at the left axillary line, and on the sixth rib in the

mid-clavicular line. To reduce possible violations of stationarity, the ECG acquisition procedure was standardized and kept consistent for all recordings (Berntson et al., 1997). The study was reviewed and approved by the St. Justine Hospital Institutional Review Board (#2040).

2.2. Procedure

2.2.1. Data processing procedure—ECG Holter tapes underwent identical processing procedures for each software program. Triplicate ECG data signals were derived from each of the 20 recordings. Each triplicate ECG recording was cleaned by a qualified investigator and independently auto scored with all three signal processing software programs strictly adhering to both Task Force (1996) guidelines and manufacturer specifications (described in detail below; see Table 1).

2.2.2. A/D data conversion

2.2.2.1. MARS: From the Holter tapes, ECG data files were downloaded and formatted into the MARS® Holter Analysis Workstation v.7.0 (Milwaukee, Wisconsin, USA).

2.2.2.2. MindWare and Kubios: ECG Holter tapes were converted and digitized into Waveform Audio (WAV) version using a high-grade contemporary dual capstan deck unit. WAV files were imported into shareware software for recording and editing audio files (Audacity® v.1.2; http://audacity.sourceforge.net). The speed of the audio signal was resampled and the length, pitch, and frequency were optimized to yield clear high-quality ECG signals. Then, using a 4-channel high-level interface module in the BioNex 2SLT Chassis Assembly (MindWare Technologies Ltd., Columbus, Ohio, USA) and the Biolab 3.0 data acquisition software (16-bit A/D conversion) the resampled digital data files were imported (sampled at 250 ks/s rate), converted, and formatted into MindWare (MW) files, while preserving the integrity of the signal. One set of raw MW formatted data files were imported into MindWare® HRV Scoring Module v.3.0.17 (MindWare Technologies Ltd., Columbus, Ohio, USA). A duplicate set was converted into ASCII text files and imported into Kubios® HRV v.2.0 (University of Eastern Finland, Kuopio, Finland; Niskanen et al., 2004). It is important to note that all software programs were used without applying any ad hoc custom-made routine changes (i.e., all default settings and specifications were maintained). The only exception was the adjustment of the default frequency bandwidths for LF and HF in MindWare; these were adjusted in accordance with the Task Force (1996) guidelines. Signal processing and default specifications are outlined below for each software program.

2.2.3. Data cleaning—Beat-by-beat intervals with near millisecond measurement of continuous ECG data were required for data cleaning. Missed or unidentified R-peaks by each respective program's detector algorithm were manually relabeled (refer to Table 1; data cleaning section). In conjunction with each software program's automated cleaning procedure, pre-defined cleaning guidelines adhering to the recommendations in the expert committee report were used by a trained investigator to accurately discriminate QRS complexes (Berntson et al., 1997). If an R-peak was automatically detected, but upon visual inspection was not found to be accurate, ≥2 short inter-beat-intervals were added to retain

the integrity of the heart period series. If an R-peak was not automatically detected, the following guidelines (in rank order) were applied: 1) RR interval distance from cleaned ECG recording sample was measured, 2) R-peak was estimated from remaining data points, and 3) long R-peak were split into ≥ 2 equal RR intervals (Berntson et al., 1997).

2.2.4. ECG signal processing

2.2.4.1. MARS: Signal processing specifications for detector algorithms and interpolation methods were based on default settings (refer to Table 1). Detector algorithms require at least 5 min of data to calculate HRV indices (adjustable). Beat-by-beat visual inspection of the shape, trend, and length of each QRS complex were measured and identified based on template matching and standard Marquette algorithms for QRS labeling. ECG data was sampled at various rates resulting in QRS timing at different resolutions (1024 samples/300 s) and RR filtering was automatic (manual filter available). The removal of artifacts was based on a 20% change from the previous signal as a criterion (Kleiger et al., 1987). In cases where artifacts and excluded RR intervals were automatically filtered and identified as unreadable signals, the remaining acceptable beats were used to replace the data points via cubic spline interpolation method. At least 4 acceptable R-peaks were needed in order for spline interpolation to identify the continuous function between two middle R-peaks. If there was no data in the first segment (e.g., noise), then RR interval series were interpolated from the default heart rate of 70 bpm (adjustable).

For spectral analyses, trending, interpolation rate, interpolation method, and windowing options (e.g., window width and overlapping) were based on default settings. Heart period series were linearly detrended, tapered using a Hanning window, and processed by FFT periodogram spectrum method. Time-and frequency-domain parameters were automatically calculated for each 5 min epoch across the entire data file. HRV parameters were then automatically averaged across the entire recording period.

2.2.4.2. MindWare: Signal processing specifications for detector algorithms could be manually overwritten, and included inter-beat-interval check and automated Minimum Artifact Deviation and Maximum Expected Deviation (MAD/MED) algorithm (Berntson et al., 1990). For the present study, 5 min analytical epochs and both detector algorithms were applied. R-peak detection was based on default digital low- and high-pass filters set within appropriate frequency ranges (0.05 and 35 Hz, respectively; adjustable). Frequency bandwidths were user-defined for LF (0.04–0.15 Hz) and HF (0.15–0.40 Hz). Beat-by-beat visual inspection of the shape, trend, and length of each QRS complex data was displayed on a full graphical interface. ECG signals were sampled at 1000 Hz and RR filtering was automatic (manual filter available). RR intervals that were excluded due to unreadable signals or recognition error were replaced by cubic spline interpolation and resampled at a frequency of 33.33 Hz.

Spectral analyses were performed on a series of RR intervals and were first linearly detrended using a Hanning window and processed by FFT standard power spectrum method. All time- and frequency-domain variables were automatically calculated for each 5 min

epoch and averaged across the entire recording period, except for SDANN and SDNNi, which were manually calculated using standard formulae (Task Force, 1996).

2.2.4.3. Kubios: Signal processing specifications for detector algorithms and interpolation methods were based on default settings (adjustable; refer to Table 1). Visual inspection of the beat-by-beat RR intervals were measured and identified based on template matching and proprietary algorithms. The sampling frequency was based on beat-by-beat RR intervals and automatically filtered, where RR intervals were divided into 5 min non-overlapping segments. As recommended by Kubios, based on visual inspection using the graphical interface, an artifact correction level (range from none to very strong) was selected for each date file. Each correction level applies thresholds (very low: 0.45 s, low: 0.35 s, medium: 0.25 s, strong: 0.15 s, very strong: 0.05 s) that are scaled with a heart rate of 60 beats/min. Scaling is used to adjust for heart rate changes within the recording (i.e., higher heart rate applies greater thresholds). High-pass filters on RR interval series remove all baseline changes from the data file, and from this detrended data, any beats that exceed the respective thresholds are identified as artifacts and removed (M.P. Tarvainen, personal communication, March 21, 2012). Because data cleaning is limited to this gross categorization to detect artifacts, Kubios recommends that artifact correction level should not be selected blindly, but should include manual visual inspection and verification of the correction level selected within the graphical interface. Continuous heart period series were corrected by piecewise cubic spline interpolation method at the default rate of 4 Hz (adjustable). Using a window width of 256 s (window overlap of 50%; adjustable), samples were smoothed prior to detrending, tapered using a Hanning window, and processed by the Welch's periodogram method.

2.3. Analysis plan

All data were entered and double-checked by the senior data coordinator and analyzed with IBM SPSS Statistics 20 software (SPSS, Inc., Chicago, IL). Data were kept continuous and checked for normality and linearity using boxplots and histograms. Assumptions of additivity, homoscedasticity, uncorrelated error, and random selection of participants were tested (Shrout and Fleiss, 1979).

To assess measurement fidelity across the three software programs, Intraclass Correlation Coefficients (ICC), Pearson Correlation Coefficients, and Bland–Altman statistical methods were computed. An ICC is a measure of agreement between two or more evaluation methods on the same data that allows for fixed and random effects. Data are assumed to be parametric (continuous and normally distributed). ICCs typically range from 0 to 1, but can exceed −1 or 1, which may be attributable to patterns of negative and positive correlations among the methods, limited variance in the data matrix, or no correlations among methods (Lahey et al., 1983). ICCs are categorized as very poor (0–0.2), fair (0.3–0.4), moderate (0.5–0.6), strong (0.7–0.8), or excellent (0.9–1.0; Shrout and Fleiss, 1979). ICCs are deemed advantageous over bivariate correlation coefficients as they represent the correspondence between two or more methods, and importantly, adjust for the effects of the scale of measurement. In other words, ICCs account for differences in rank order and mean differences between methods (data centered and scaled using pooled mean across methods

and standard deviation), while correlations only account for rank order differences (data centered and scaled using each method's own mean and standard deviation). Nevertheless, Pearson Correlation Coefficients were computed for comparison purposes. Analysis of variance (ANOVA) was also used to test omnibus mean differences of the HRV parameters, followed by contrasts using paired samples t-tests.

The Bland–Altman method is used to graphically display the degree of agreement between two techniques on a continuous variable and to assess possible constant and proportional biases (Bland and Altman, 1986, 2003). The differences in the measurements are plotted against the mean values of these measurements. If 95% of the differences fall within the limits of agreement (1 SD) there is no systematic variation across programs (Bland and Altman, 1986, 2003). To detect constant bias (i.e., the average discrepancy between methods of measurements), the mean bias and limits of agreement are used and should be close to zero. To detect proportional bias, visual inspection of the plotted graphs is commonly used; however, standardized β values can be used to test whether the slope is significantly different than zero (i.e., when mean values are regressed onto mean differences).

3. Results

The average length of the 20 ECG recordings was 131 min (SD= 46). All ECG recordings were inspected manually to review peak detection and to identify and remove artifacts. Manual editing took approximately 25 min per ECG recording. Recordings were found to be of excellent quality; over 90% of data were analyzable, artifact time did not exceed 1500 s (5.2%), and no recordings were found to exceed 20% noise or ectopic beats.

ICCs were computed to compare the fidelity of HRV scoring across the software programs (see Table 2). Among the time-domain indices, there was strong to excellent correspondence across all software programs for SDNN (ICC_{avg} =0.96; r_{avg} =0.97), SDANN (ICC_{avg} =0.93; r_{avg} =0.88), SDNNi (ICC_{avg} =0.96; r_{avg} =0.97), rMSSD (ICC_{avg} =0.80; r_{avg} =0.93), and pNN50 (ICC_{avg} =0.98; r_{avg} =0.99). Among the frequency-domain indices, there was excellent correspondence across all software programs for LF (ICC_{avg}=0.90; r_{avg} =0.94), HF (ICC_{avg}=0.91; r_{avg} =0.96), and LF/HF ratio (ICC_{avg} =0.95; r_{avg} =0.93). However, VLF exhibited poor correspondence (ICC_{avg} =0.19); these findings may be largely attributable to the significant mean level differences observed across software programs (see Table 3). Pearson coefficients revealed moderate correlations for VLF when mean level differences are not considered (r_{avg} =0.83).

Bland–Altman plots and analyses were conducted to assess measurement fidelity for each HRV parameter paired by software programs (30 plots not depicted for parsimony). For each HRV parameter, the differences between each of the paired software programs were plotted against the average values of these measurements. Consistent with the recommendations outlined by Bland and Altman (1986, 2003), data were log-transformed prior to the calculation of limits of agreement when heteroscedasticity was present. There was no evidence of constant or proportional biases for any of the time-domain variables: SDNN ($Bias_{avg}$=0.02, [Limits of $Agreement_{avg}$=−0.03, 0.08]; β_{avg}= −0.07), SDANN ($Bias_{avg}$=0.04, [−0.05, 0.14]; β_{avg}=0.05), SDNNi ($Bias_{avg}$= 0.03, [−0.06, 0.13]; β_{avg}=−0.16), rMSSD

($Bias_{avg}$=0.09, [−0.00, 0.19]; β_{avg}=0.07), and pNN50 ($Bias_{avg}$=0.07, [−0.09, 0.25]; β_{avg}= − 0.06). Similarly, no constant or proportional biases were observed for the frequency-domain variables: VLF ($Bias_{avg}$ =0.70, [0.43, 0.96]; β_{avg} = −0.00), LF ($Bias_{avg}$ =0.10, [−0.02, 0.22]; β_{avg} = −0.19), HF ($Bias_{avg}$ =0.13, [−0.01, 0.29]; β_{avg} =0.22), and LF:HF ratio ($Bias_{avg}$ =0.10, [−0.02, 0.22]; β_{avg} = −0.11). Altogether, the results from the ICCs and Bland–Altman analyses were congruent.

4. Discussion

Recent advances in the automated analyses of HRV offers an accessible and unique approach for quantifying the effects of sympathetic and parasympathetic branches of the ANS. Despite evidence of the reliability of HRV parameters across different recording devices, measurement protocols, and maneuvers (Dietrich et al., 2010; Faulkner et al., 2003; Pinna et al., 2007; Sandercock, Shelton and Brodie, 2004, 2005, 2003), there is no available information on the fidelity of commercially available signal processing software programs currently in use (Jung et al., 1996). The aim of the present study was to evaluate the measurement fidelity of HRV indices derived from three commonly used signal processing software programs.

Following stringent standardization (i.e., data collection, processing, cleaning), excellent measurement fidelity for time-domain variables (e.g., SDNN, SDANN, SDNNi, rMSSD, pNN50) was observed across programs. Excellent correspondence was also observed for LF, HF, and LF/HF ratio. Poor correspondence was found for VLF; however, examination of the Pearson correlation indicates a moderate association across software programs. The excellent comparability for HRV variables is likely attributable to similar signal processing techniques and pivotal user-defined specifications across software programs (i.e., R-peak detection algorithm, identical analytical epoch length). For instance, the use of algorithms parallel to the Pan-Tompkins for the recognition of QRS complexes was apparent across all software programs (Pan and Tompkins, 1985). As such, the ECG signal is passed through an automated low- and high-pass filter to remove noise. After filtering, the signal passes through derivative (to obtain QRS slope), squaring (to emphasize higher frequencies), and window integration phases (to identify waveform patterns), where lastly, a threshold method is applied and R-peaks are detected. As for the frequency-domain variables, windowing options (i.e., width and overlap) and frequency bandwidths must also be taken into consideration (Task Force, 1996). In the present study, all software programs applied linear detrending method, cubic spline interpolation, with similar windowing (Hamming and Hanning) and spectrum methods (Periodogram and Welch's periodogram).

User-defined data reduction decisions can have significant implications on the automatic analysis of HRV parameters. Short analytical epochs (e.g., 1 min) and recording durations (<18 h) may fail to capture the full spectrum of components or underlying circadian rhythms (Massin et al., 2000; Task Force, 1996). For example, the lowest frequency that can be assessed with 1 min is 0.016 Hz (G. Berntson, personal communication, December 15, 2011), indicating that it does not quantify the full spectrum of VLF components. Thus, to capture data at the lowest frequency, larger analytical epoch durations must be chosen (e.g., 3 to 5 min; Task Force, 1996). Further, the established physiological components and

frequency bandwidth ranges are less well-defined for VLF, as compared to HF and LF (Berntson et al., 1994; Cacioppo et al., 1994). Analytical epoch length, recording durations, and frequency bands should be consistent when making comparisons of HRV.

Given that technical specifications for data cleaning vary across programs, it is essential to know whether programs allow for manual inspection (i.e., some permit simultaneous automatic and manual cleaning and editing decisions). For example, MindWare offers users much flexibility to visually inspect and adjust RR fiducial points and identify important event markers (e.g., during tasks). In contrast, Kubios suggests visually inspecting data and applying an automated artifact correction based on gross categorization levels (e.g., low). Given the sensitivity of certain HRV parameters (e.g., rMSSD; Salo et al., 2001), the level of gross artifact correction may be appropriate for some variables, while less appropriate for others. Taken together, these specific user-defined decisions likely account for the exceptional correspondence across software programs.

The present study yields original findings indicating the robust comparability for HRV across commonly used signal processing programs. While proprietary detector and interpolation algorithms are typically set, the excellent correspondence across software programs is largely attributable to seemingly nuanced, yet significant decisions. These include decisions related to the modification of particular user-defined and default settings (e.g., analytical epoch duration, frequency-bandwidths), use of cleaning tools (e.g., selection of appropriate artifact correction level), and inherent procedures in each software program (e.g., removing partial inter-beat intervals prior to data analysis).

Prior to selecting signal processing software, the conceptualization and understanding of HRV physiological indices is imperative. There is growing interest and advancements using neuroimaging techniques (e.g., functional magnetic resonance imaging) to better understand neurobiological (brain–body) interactions (c.f., Gianaros and Sheu, 2009; Gianaros et al., 2004). For example, HF has been associated with activity within the ventral anterior cingulate (Matthews et al., 2004), posterior cingulate cortex (O'Connor et al., 2007), amygdala, periaqueductal gray, and the hypothalamus in response to somatosensory stimuli (Gray et al., 2009) and isometric exercise (Napadow et al., 2008). Given the evidence of an association between the brain and the ANS (i.e., parasympathetic and sympathetic activity), these promising research directions underscore the importance of purposeful and informed selection of HRV parameters. Consider, if the research question centers around assessing parasympathetic activity, it is necessary to select HRV parameters that validly reflect this activity in the ANS (e.g., HF, pNN50, rMSSD; Task Force, 1996). This in turn will directly impact decisions related to methodological design and measurement issues, including the recommended recording length to capture parasympathetic activity (e.g., 1 min), and an effort to minimize non-stationarity across conditions and participants, particularly for frequency-domain variables (Task Force, 1996). Other decisions may include whether recordings will be partitioned by task or interval (e.g., baseline vs. task, sleep vs. wake state). Similar issues were eloquently raised in a thorough review by Nunan et al. (2010) investigating normative HRV values from short-term recordings in healthy adults. Taking these pivotal methodological decisions into consideration will facilitate comprehensive systematic comparisons across studies and further advance the field.

4.1. Post-hoc observations

4.1.1. Kubios—Several researchers report using an alternate strategy to clean data prior to using Kubios by deleting aberrant inter-beat intervals less than 300 and greater than 1200 ms (c.f., Capa et al., 2011; Li et al., 2009; Rodríguez-Colón et al., 2011; Timonen et al., 2006). Data were re-analyzed with Kubios after applying this commonly reported data cleaning strategy. Post-hoc analyses revealed no significant differences across software programs for both time- and frequency-domain variables when this data cleaning strategy was applied (data not shown for parsimony).

4.2. Strengths and limitations

The first limitation of the present study was the use of short-rather than long-term recordings (i.e., 3 vs. 24 h). However, many studies typically record for similarly short durations. In keeping with the recommendations by the Task Force (1996), the present study adhered to a strict protocol for the acquisition, recording, collection, cleaning, and analyses of the data under standardized settings to minimize measurement error.

The second limitation was the use of only three software programs for comparison. These programs were purposely selected due to their ubiquitous use within clinical and research settings among psycho-physiologists, cardiologists, and general researchers. Nevertheless, it is important to recognize there are additional commercially available as well as investigator-created software programs; however, their inclusion was beyond the scope of the present study. Future comparisons should be conducted using other software programs.

The third limitation was the assessment of only time- and frequency-domain variables. Geometric (e.g., triangular shapes of Lorenz plots) and nonlinear methods (e.g., detrended fluctuation analysis, approximate entropy) can also be used to analyze HRV (Pincus, 1995; Porta et al., 2001, 2007; Richman and Moorman, 2000; Task Force, 1996; Voss et al., 2009). However, these methods largely depend on the precision of equipment (i.e., obtain appropriate number of RR intervals), recording length (i.e., preferably 24 h for geometric methods), and capability of these advanced analyses in software programs. Time- and frequency-domain variables are traditional HRV parameters reported in the majority of studies; thus, the comparability of these specific parameters was deemed particularly important to inform future comparisons and syntheses across published studies (Task Force, 1996).

Lastly, all ECG recordings were derived from a Holter monitor manufactured by GE Marquette, the same manufacturer of MARS software program. However, it is unlikely that having a common manufacturer created any bias for the MARS software analyses. In fact, a major strength of the present study was the use of identical ECG recordings in triplicate for the three software programs. In other words, each software program analyzed the exact same ECG data. Thus, these findings are generalizable to the scenario quite common in research and clinical settings when hardware and software manufacturers differ.

4.3. Recommended strategies

Although there are an increasing number of studies investigating HRV, the methodological, measurement, and technical specifications are not consistently applied in the field. These discrepancies add confusion to the interpretation of HRV and hinder advancement in the field because findings cannot be synthesized. Hence, to maximize measurement fidelity researchers must be cognizant of these subtle, yet pivotal fine details when using software programs. Two recommended strategies are provided.

4.3.1. Equipment and software specifications—Differences across user-defined choices and specifications of software programs may contribute to HRV discrepancies across studies. Researchers should report specific information about the recording equipment, signal (pre)processing software, software applications, and features selected (e.g., sampling rate of 250–500 Hz or higher, RR interval filter characteristics, R-peak detection and interpolation algorithms). Further, if frequency-domain variables are analyzed, additional information on the spectral decomposition method, spectral windowing, window overlap, and the defined range of frequency bandwidths should be specified.

4.3.2. Data reduction and cleaning—Data reduction and cleaning decisions prior to HRV analysis (either by default or adjustable settings) should be explained and justified. For example, because the removal of erroneous beats or the unintentional removal of normal beats may affect the analysis and the comparability of HRV parameters (Berntson et al., 1997; Berntson and Stowell, 1998; Xia et al., 1993), the rationale for any exclusion criteria should be clearly stated. Furthermore, to facilitate systematic comparisons and synthesis of data, it is important to provide complete information on data reduction decisions. These include justification for how the data were segmented or partitioned for aggregating (e.g., conditions, tasks, control vs. clinical groups), cleaning (e.g., duration of analytical epochs), and analyzing (e.g., night vs. day). Complex study designs (e.g., multiple discreet intervals) may warrant use of software that permits greater flexibility for user-specifications and manual cleaning (i.e., Mindware). Regardless of what equipment or software is used, movement artifacts, technical failure, or poor data quality can seriously contaminate the integrity of the data. Despite the crucial task of manually cleaning data, specific procedures and decision rules are rarely reported. Basic information on the RR interval error identification, removal, criteria (e.g., thresholds), and correction procedures should be provided.

4.4. Future research

Future studies should assess the measurement fidelity of time- and frequency-domain HRV variables with longer recordings (e.g., 24 h), under differing conditions (e.g., day vs. night), and in response to standardized challenges (e.g., stress testing, cold pressor reactivity). Additional geometric methods (i.e., HRV triangular index) should also be considered. Further, comparisons could be made for HRV parameters derived from different recording hardware and then analyzed with different software programs, as this would be a more ecologically valid reflection of the diverse practices across the research field. The contribution of the present study highlights the importance of providing sufficient detail about the signal acquisition hardware, the signal processing software, and the overall

procedures used to derive HRV variables. Lastly, given that guidelines to specify standard definitions of HRV terms and measurement methodology were published almost two decades ago (e.g., Task Force, 1996; Berntson et al., 1997), there is merit in the proposal of updating the critical considerations in HRV analyses (e.g., Nunan et al., 2010).

4.5. Conclusion

The present study demonstrated that stringent decisions and specifications for subtle details are instrumental in the acquisition of excellent measurement fidelity across three commonly used HRV signal processing software programs. Specifically, signal processing, data cleaning, analysis, and interpretation specifications must be meticulously selected to enhance the precision of HRV data and should not be underestimated. Given the significance and value of comparing and synthesizing results across studies, it is crucial for researchers to understand and accurately report the technical specifications applied for HRV analyses.

Acknowledgments

Sincere thanks to the developers, technicians, and customer support staff of the investigated software programs for their extensive help, clarification, and detailed information provided (Kubios: Mika P. Tarvainen; Mindware: Gary Bernston, Marty Gillman, Greg Norman, Doug Schiefer, Dave Lozano, Gene Barbanera, Sean Canavan; MARS/GE Marquette Medical Systems: Jason Castillo, Pascal Langlois, Marilyn Manders). This work was made possible through funding support from the Canadian Institute of Health Research (CIHR) New Investigator Award (J. J. McGrath), Fonds de la Recherche en Santé du Québec Doctoral Fellowship (FRSQ) and Health Professional Student Research Award (D. Jarrin), and FRSQ Senior Clinical Scientist Award (P. Poirier). The Holter recordings were obtained from the QUALITY Cohort, which is funded by CIHR, FRSQ, and the Heart and Stroke Foundation of Canada. The QUALITY Cohort was conducted by members of Team PRODIGY, an inter-university research team including Université de Montréal, Concordia University, Université Laval, McGill University, and INRS-Institut Armand Frappier.

Disclosure: There are no conflicts of interest to disclose. No software companies provided any compensation or financial support. All software was procured through standard procedures; no complimentary software access was provided. The use of the selected hardware and software programs does not necessarily constitute or imply their endorsement or agreement with the results of this study. The software companies were not involved in the study design, data analyses, data interpretation or manuscript writing and submission processes.

References

Akselrod S, Gordon D, Ubel FA, Shannon DC, Barger AC, Cohen RJ. Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. Science. 1981; 213:220–222. [PubMed: 6166045]

Bailey JJ, Horton M, Itscoitz SB. A method for evaluating computer programs for electrocardiographic interpretation. III. Reproducibility testing and the sources of program errors. Circulation. 1974; 50:88–93. [PubMed: 4276020]

Bailey JJ, Berson AS, Garson A Jr, Horan LG, Macfarlane PW, Mortara DW, Zywietz C. Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing: a report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology American Heart Association. Circulation. 1990; 81:730–739. [PubMed: 2297875]

Berntson GG, Stowell JR. ECG artifacts and heart period variability: don't miss a beat! Psychophysiology. 1998; 35:127–132. [PubMed: 9499713]

Berntson GG, Quigley KS, Jang JF, Boysen ST. An approach to artifact identification: application to heart period data. Psychophysiology. 1990; 27:586–598. [PubMed: 2274622]

Berntson GG, Cacioppo JT, Quigley KS. Autonomic determinism: the modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. Psychological Review. 1991; 98:459–487. [PubMed: 1660159]

Berntson GG, Cacioppo JT, Quigley KS. Autonomic cardiac control. I. Estimation and validation from pharmacological blockades. Psychophysiology. 1994; 31:572–585. [PubMed: 7846218]

Berntson GG, Bigger JT, Eckberg DL, Grossman P, Kaufmann PG, Malik M, Nagaraja HN, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997; 34:623–648. [PubMed: 9401419]

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]

Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound in Obstetrics & Gynecology. 2003; 22:85–93. [PubMed: 12858311]

Bloomfield, P. Fourier analysis of time series: an introduction. John Wiley; New York: 1976.

Bonner RE, Schwetman HD. Computer diagnosis of electrocardiograms II: a computer program for EKG measurements. Computers and Biomedical Research. 1968; 1:366–386. [PubMed: 5696978]

Bonner RE, Crevasse L, Ferrer MI, Greenfield JC Jr. A new computer program for analysis of scalar electrocardiograms. Computers and Biomedical Research. 1972; 5:629–653. [PubMed: 4264999]

Cacioppo JT, Berntson GG, Binkley PF, Quigley KS, Uchino BN, Fieldston A. Autonomic cardiac control. II. Basal response, noninvasice indices, and autonomic space as revealed by autonomic blockades. Psychophysiology. 1994; 31:586–598. [PubMed: 7846219]

Camm AJ, Pratt CM, Schwartz PJ, AL-Khalidi HR, Spyt MJ, Holroyde MJ, Karam R, et al. Mortality in patients after a recent myocardial infarction: a randomized, placebo-controlled trial of azimilide using heart rate variability for risk stratification. Circulation. 2001; 109:990–996.

Capa RL, Cleeremans A, Bustin GM, Hansenne M. Long-lasting effect of subliminal processes on cardiovascular responses and performance. International Journal of Psychophysiology. 2011; 81:22–30. [PubMed: 21515314]

Dietrich A, Rosmalen JGM, Althaus M, van Roon AM, Mulder LJM, Minderaa RB, Oldehinkel AJ, Riese H. Reproducibility of heart rate variability and baroreflex sensitivity measurements in children. Biological Psychology. 2010; 85:71–78. [PubMed: 20553793]

Faulkner MS, Hathaway D, Tolley B. Cardiovascular autonomic function in healthy adolescents. Heart & Lung. 2003; 32:10–22. [PubMed: 12571544]

Ferri R, Curzi-Dascalova L, Arzimanoglou A, Bourgeois M, Beaud C, Lahorgue Nunes M, Elia M, et al. Heart rate variability during sleep in children with partial epilepsy. Journal of Sleep Research. 2002; 11:153–160. [PubMed: 12028480]

Friedman BH. An autonomic flexibility-neurovisceral integration model of anxiety and cardiac vagal tone. Biological Psychology. 2007; 74:185–199. [PubMed: 17069959]

Friedman BH, Thayer JF. Anxiety and autonomic flexibility: a cardiovascular approach. Biological Psychology. 1998; 49:303–323. [PubMed: 9858059]

Gerritsen J, Dekker JM, TenVoorde BJ, TenVoorde BJ, Kostense PJ, Heine RJ, Bouter LM, et al. Impaired autonomic function is associated with increased mortality, especially in subjects with diabetes, hypertension, or a history of cardiovascular disease: The Hoorn Study. Diabetes Care. 2001; 24:1793–1798. [PubMed: 11574444]

Gianaros PJ, Sheu LK. A review of neuroimaging studies of stressor-evoked blood pressure reactivity: emerging evidence for a brain-body pathway to coronary heart disease risk. NeuroImage. 2009; 47:922–936. [PubMed: 19410652]

Gianaros PJ, Van Der Veen FM, Jennings JR. Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working-memory tasks: implications for the cortical and subcortical regulation of cardiac autonomic activity. Psychophysiology. 2004; 41:521–530. [PubMed: 15189475]

Gray MA, Minati L, Harrison NA, Gianaros PJ, Napadow V, Critchley HD. Physiological recordings: basic concepts and implementation during functional magnetic resonance imaging. NeuroImage. 2009; 47:1105–1115. [PubMed: 19460445]

Harris FJ. On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE. 1978; 66:51–83.

Hemingway H, Shipley M, Brunner E, Britton A, Malik M, Marmot M. Does autonomic function link social position to coronary risk? The Whitehall II study. Circulation. 2005; 111:3071–3077. [PubMed: 15939818]

Jezewski J, Kupka T, Horoba K. Extraction of fetal heart rate signal as time event series from evenly sampled data acquired using Doppler ultrasound technique. IEEE Transactions on Biomedical Engineering. 2008; 55:805–810. [PubMed: 18270022]

Jung J, Heisel A, Tscholl D, Fries R, Schieffer H, Ozbek C. Assessment of heart rate variability by using different commercially available systems. The American Journal of Cardiology. 1996; 78:118–120. [PubMed: 8712103]

Kim KK, Lim YG, Kim JS, Park KS. Effect of missing RR-interval data on heart rate variability analysis in the time domain. Physiological Measurement. 2007; 28:1485–1494. [PubMed: 18057513]

Kim KK, Kim JS, Lim YG, Park KS. The effect of missing RR-interval data on heart rate variability analysis in the frequency domain. Physiological Measurement. 2009; 30:1039–1050. [PubMed: 19713596]

Kitney, R. An analysis of the thermoregulatory influences on heart-rate variability. Kitney, RI., Rompelman, O., editors. Clarendon Press; Oxford, England: 1980.

Kleiger RE, Miller JP, Bigger JT Jr, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. Journal of the American College of Cardiology. 1987; 59:256–262.

Kligfield P, Gettes L, Bailey JJ, Childers R, Deal BJ, Hancock W, van Herpen G, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. Journal of the American College of Cardiology. 2007; 49:1109–1127. [PubMed: 17349896]

Köhler BU, Hennig C, Orglmeister R. The principles of software QRS detection. IEEE Engineering in Medicine and Biology Magazine. 2003; 21:42–57.

Lahey MA, Downey RG, Saal FE. Intraclass correlations: there's more than meets the eye. Psychological Bulletin. 1983; 93:586–595.

Lahiri MK, Kannankeril PJ, Goldberger JJ. Assessment of autonomic function in cardiovascular disease: physiological basis and prognostic implications. Journal of the American College of Cardiology. 2008; 51:1725–1733. [PubMed: 18452777]

Lambert, M., Van Hulst, A., O'Loughlin, J., Tremblay, A., Barnett, TA., Charron, H., Drapeau, V., et al. Cohort profile: the Quebec adipose and lifestyle investigation in youth cohort. International Journal of Epidemiology. 2011. http://dx.doi.org/10.1093/ije/dyr111 (Advance online publication http://ije.oxfordjournals.org/content/early/2011/07/23/ije.dyr111.short?rss=1)

Li Z, Snieder H, Su S, Ding X, Thayer JF, Treiber FA, Wang X. A longitudinal study in youth of heart rate variability at rest and in response to stress. International Journal of Psychophysiology. 2009; 73:212–217. [PubMed: 19285108]

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; 51:3524–3531. [PubMed: 12453910]

Lindmark S, Wiklund U, Bjerle P, Eriksson JW. Does the autonomic nervous system play a role in the development of insulin resistance? A study on heart rate variability in first-degree relatives of Type 2 diabetes patients and control subjects. Diabetic Medicine. 2003; 20:399–405. [PubMed: 12752490]

Lippman N, Stein KM, Lerman BB. Differential therapeutic responses of patients with isoproterenol-dependent and isoproterenol-independent vasodepressor syncope. American Heart Journal. 1994; 128:1110–1116. [PubMed: 7985591]

Lorena SL, Figueiredo MJ, Almeida JR, Mesquita MA. Autonomic function in patients with functional dyspepsia assessed by 24-hour heart rate variability. Digestive Diseases and Sciences. 2002; 47:27–31. [PubMed: 11837729]

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 Tech. 2009; 33:634–641.

Malik, M., Camm, AJ. Heart Rate Variability. Futura Publishing Company; Armonk, NY: 1995.

Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. Circulation. 1991; 84:482–492. [PubMed: 1860193]

Mase A, Nagae D, Ito N, Komada Y. Non-contact and non-invasive stress evaluation using microwave reflectometry (Abstract). Electromagnetics in Advanced Applications (ICEAA). 2010:144.

Masi CM, Hawkley LC, Rickett EM, Cacioppo JT. Respiratory sinus arrhythmia and diseases of aging: obesity, diabetes mellitus, and hypertension. Biological Psychology. 2007; 74:212–223. [PubMed: 17034928]

Massin MM, Maeyns K, Withofs N, Ravet F, Gérard P. Circadian rhythm of heart rate and heart rate variability. Archives of Disease in Childhood. 2000; 83:179–182. [PubMed: 10906034]

Matthews SC, Paulus MP, Simmons AN, Nelesen RA, Dimsdale JE. Functional subdivisions within anterior cingulate cortex and their relationship to autonomic nervous system function. NeuroImage. 2004; 22:1151–1156. [PubMed: 15219587]

Mazurak N, Stein J, Kipphan S, Muth ER, Teufel M, Zipfel S, Enck P. Heart rate variability in anorexia nervosa and the irritable bowel syndrome. Neurogastroenterology and Motility. 2011; 23:e470–e478. [PubMed: 21917084]

Montano, N., Tobaldini, E., Porta, A. The autonomic nervous system. In: Chouker, A., editor. Stress Challenges and Immunity in Space. Springer; Heidelberg, Germany: 2012.

Mortara, DW. Computers in Cardiology. Institute of Electrical and Electronics Engineers; New York, NY: 1977. Digital filters for ECG signals; p. 511-514.

Napadow V, Dhond R, Conti G, Makris N, Brown EN, Barbieri R. Brain correlates of autonomic modulation: combining heart rate variability with fMRI. NeuroImage. 2008; 42:169–177. [PubMed: 18524629]

Niskanen JP, Tarvainen MP, Rantaaho PO, Karjalainen PA. Software for advanced HRV analysis. Computer Methods and Programs in Biomedicine. 2004; 76:73–81. [PubMed: 15313543]

Nugent AC, Bain EE, Thayer JF, Sollers JJ, Drevets WC. Heart rate variability during motor and cognitive tasks in females with major depressive disorder. Psychiatry Research. 2011; 191:1–8. [PubMed: 21129936]

Nunan D, Sandercock GRH, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. Pacing and Clinical Electrophysiology : PACE. 2010; 33:1407–1417. [PubMed: 20663071]

O'Connor MF, Gundel H, McRae K, Lane RD. Baseline vagal tone predicts BOLD response during elicitation of grief. Neuropsychopharmacology. 2007; 32:2184–2189. [PubMed: 17299507]

Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering. 1985; 32:230–236. [PubMed: 3997178]

Parati G, Saul JP, Di Rienzo M, Mancia G. Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation: a critical appraisal. Hypertension. 1995; 25:1276–1286. [PubMed: 7768574]

Pincus SM. Approximate entropy (ApEn) as a complexity measure. Chaos. 1995; 5:110–118. [PubMed: 12780163]

Pinna G, Maestri R, Torunski A, Danilowicz-Szymanowicz L, Szwoch M, La Rovere MT, Raczak G. Heart rate variability measures: a fresh look at reliability. Clinical Science (London, England). 2007; 113:131–140.

Pipberger HV, Stallman FW, Berson AS. Automatic analysis of the P-QRS-T complex of the electrocardiogram by digital computer. Annals of Internal Medicine. 1962; 57:776–787. [PubMed: 13943712]

Pomeranz B, Macaulay RJB, Caudill MA, Kutz I, Adam D, Gordon D, Kilborn KM, et al. Assessment of autonomic function in humans by heart rate spectral analysis. The American Journal of Physiology. 1985; 248:H151–H153. [PubMed: 3970172]

Porta A, Guzzetti S, Montano N, Furlan R, Pagani M, Malliani A, Cerutti S. Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. IEEE Transactions on Biomedical Engineering. 2001; 48:1282–1291. [PubMed: 11686627]

Porta A, Guzzetti S, Furlan R, Gnecchi-Ruscone T, Montano N, Malliani A. Complexity and nonlinearity in short-term heart period variability: comparison of methods based on local nonlinear

prediction. IEEE Transactions on Biomedical Engineering. 2007; 54:94–106. [PubMed: 17260860]

Pryor TA, Russell R, Budkin A, Price WG. Electrocardiographic interpretation by computer. Computers and Biomedical Research. 1969; 2:537–548. [PubMed: 4904456]

Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart and Circulatory Physiology. 2000; 278:H2039–H2049. [PubMed: 10843903]

Rodríguez-Colón SM, Bixler EO, Li X, Vgontzas AN, Liao D. Obesity is associated with impaired cardiac autonomic modulation in children. International Journal of Pediatric Obesity. 2011; 6:128–134. [PubMed: 20919806]

Romhilt DW, Estes EH Jr. A point-score system for the ECG diagnosis of left ventricular hypertrophy. American Heart Journal. 1968; 75:752–758. [PubMed: 4231231]

Salo MA, Huikuri HV, Seppanen T. Ectopic beats in heart rate variabilityanalysis: effects of editing on time and frequency domain measures. Annals of Noninvasive Electrocardiology. 2001; 6:5–17. [PubMed: 11174857]

Sandercock GRH, Shelton C, Brodie DA. Heart rate variability instrumentation: agreement and reliability. The Journal of Physiology. 2003; 552P:C34.

Sandercock GR, Bromley P, Brodie DA. Reliability of three commercially available heart rate variability instruments using short-term (5-min) recordings. Clinical Physiology and Functional Imaging. 2004; 24:359–367. [PubMed: 15522045]

Sandercock GRH, Bromley PD, Brodie DA. The reliability of short-term measurements of heart rate variability. International Journal of Cardiology. 2005; 103:238–247. [PubMed: 16098384]

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428. [PubMed: 18839484]

Spiers JP, Silke B, McDermott U, Shanks RG, Harron DW. Time and frequency domain assessment of heart rate variability: a theoretical and clinical appreciation. Clinical Autonomic Research. 1993; 3:145–158. [PubMed: 8324377]

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation. 1996; 93:1043–1065. [PubMed: 8598068]

Taylor JA, Carr DL, Myers CW, Eckberg DL. Mechanisms underlying very-low-frequency RR-interval oscillations in humans. Circulation. 1998; 98:547–555. [PubMed: 9714112]

Thayer JF, Lane RD. The role of vagal function in the risk for cardiovascular disease and mortality. Biological Psychology. 2007; 74:224–242. [PubMed: 17182165]

Timonen KL, Vanninen E, de Hartog J, Ibald-Mulli A, Brunekreef B, Gold DR, Heinrich J, Hoek G, Lanki T, Peters A, Tarkiainen T, Tiittanen P, Kreyling W, Pekkanen J. Effects of ultrafine and fine particulate and gaseous air pollution on cardiac autonomic control in subjects with coronary artery disease: the ULTRA study. Journal of Exposure Science & Environmental Epidemiology. 2006; 16:332–341. [PubMed: 16205787]

van Bemmel JH, Zywietz C, Kors JA. Signal analysis for ECG interpretation. Methods of Information in Medicine. 1990; 29:317–329. [PubMed: 2233378]

Voss A, Schulz S, Schroeder R, Baumert M, Caminal P. Methods derived from nonlinear dynamics for analysing heart rate variability. Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences. 2009; 367:277–296.

Welch PD. The use of fast Fourier transforms for the estimation of power spectra: a method based on time averaging over short modified periodograms. IEEE Transactions on Audio and Electroacoustics. 1967; 15:70–73.

Xia R, Odemuyiwa O, Gill J, Malik M, Camm AJ. Influence of recognition errors of computerised analysis of 24-hour electrocardiograms on the measurement of spectral components of heart rate variability. International Journal of BioMedical Computing. 1993; 32:223–235. [PubMed: 7685743]

Table 1

System-dependent specifications across signal processing software programs.

	MARS (GE Marquette)	MindWare	Kubios HRV	Task Force
• Version	–MARS Holter Analysis Workstation v7	–HRV v3.0.17	–Kubios HRV v2.0	
Signal acquisition and conversion				
• Import options	–Raw ECG signals	–Raw ECG signals, BIOPAC (.acq), Mindware (.MW)	–Only RR Intervals	
• Input files	–MARS software	–Mindware format (.MW)	–ASCII	
• A/D resolution	–Not reported	–16 bit	–Offiine analysis program	
Signal preprocessing				
• Preprocessing	–Manual visual inspection	–Manual visual inspection	–Recommends manual and visual inspection prior to using program	
• Sampling frequency	–125 Hz	–1000 Hz	–Offiine analysis program	–Optimal 250–500 Hz or higher
• R-peak detection	–Template matching (cross-correlation for upcoming signal with templates already formed)	–Low-pass and high-pass filters for raw ECG data (adjustable)	–Proprietary algorithm (akin to Pan-Tompkins)	–Use well-tested algorithm (e.g., template, cross-correlation, derivative plus threshold)
• RR interval filtering and interpolation	–Automated and manual filtering	–Automated and manual filtering	–Automated filtering only	–LF cutoff=0.05 Hz –HF cutoff =150 Hz
• Detrending or autoregressive algorithms	–Linear trend fit to FFT input of 600 s window –Middle 5 min detrended	–Linearly detrended	–Smoothness priors detrending –AR model order: 16 –None, 1st–3rd order	
• Resampling or interpolation rate	–5 min RR intervals –Spline model to interpolate to 1024 evenly spaced data	–Resampled at frequency based on 200 bpm/60×10 or 33.33 Hz	–4 Hz (default; adjustable)	–At least 512 but preferably 1024 samples for 5 min recordings
• Interpolation method	–Cubic spline interpolation (Discrete Event Series; DES)	–Cubic spline interpolation	–Cubic spline interpolation	–Regularly sampled interpolation of DES with (non)parametric methods
• Window width	–300 s	–Not reported	–256 s (default; adjustable)	
• Window overlap	–80% re-sampled	–Not reported	–50% re-sampled (adjustable)	
• Windowing	–Hanning –Spectral coefficients scaled to account for attenuation of signal energy due to window	–Hamming	–Hanning –Frequency-domain: 256 points/Hz (adjustable)	–Hanning/Hamming
• Spectrum method	–Periodogram	–Power spectrum	–Welch's periodogram	
Data cleaning				
• Artifact detection and handling	–Manual handling artifacts	–Manual handling artifacts	–Artifact correction levels:	–Interpolation on preceding or
	–GE Marquette® algorithms	–Dual ECG artifact detection algorithms, MAD/MED, IBI check (Berntson et al., 1990)	(none, very low, low, medium, strong, very strong; custom)	successive beats on HRV signal or autocorrelation function
HRV indices automatic analyses				
• Interval calculations	–Calculated for entire recording period (adjustable)	–Calculated for each epoch and averaged across entire	–Calculated for entire recording period	

	MARS (GE Marquette)	MindWare	Kubios HRV	Task Force
		recording period (adjustable)		
• Time domain indices	–Mean RR (ms) –SDNN (ms) –SDANN (ms) –SDNNi (ms) –rMSSD (ms) –NN50 (counts) –pNN50 (%)	–Mean RR (ms) –SDNN (ms) –rMSSD (ms) –NN50 (counts) –pNN50 (%)	–Mean RR (ms) –SDNN (ms) –SDANN (ms) –SDNNi (ms) –rMSSD (ms) –NN50 (counts) –pNN50 (%)	–SDNN (ms) –SDANN (ms) –SDNNi (ms) –rMSSD (ms) –NN50 (counts) –pNN50 (%)
• Frequency bands	–VLF (0.0033–0.04 Hz) –LF (0.0400–0.15 Hz) –HF (0.1500–0.4 Hz)	–VLF (0.0030–0.0400 Hz) –LF (0.0400–0.1500 Hz) –HF (0.1500–0.4000 Hz)	–VLF (0.00–0.04 Hz) –LF (0.04–0.15 Hz) –HF (0.15–0.4 Hz)	–VLF (0.00–0.04 Hz) –LF (0.04–0.15 Hz) –HF (0.15–0.4 Hz)
• Units	–Hz, ms^2	–Hz, ms^2	–Hz, ms^2, %, n.u.	–Hz, ms^2, %, n.u.
• Additional output	–# Rs detected, Ventricular and Supraventricular beats (<1%)	–# Rs detected, RSA, First ECG R time	–Geometric parameters (RR triangular index, TINN), Poincare Plot (SD1, SD2)	
• Export options	–PDF	–ASCII	–PDF, Matlab MAT-file, ASCII	

Note. Information in the table was derived from product support manuals; not all information was reported.

Table 2

Measurement fidelity for heart rate variability parameters across software programs.

	MARS vs. MindWare		MARS vs. Kubios		MindWare vs. Kubios	
	ICC (95%CI)	*r*	*ICC (95%CI)*	*r*	*ICC (95%CI)*	*r*
Mean RR (ms)	0.98 (0.93, 0.99)	0.96**	0.98 (0.94, 0.99)	0.96**	1.00 (0.99, 1.00)	1.00**
Time-domain						
SDNN (ms)	0.93 (0.80, 0.98)	0.94**	0.97 (0.93, 0.99)	0.97**	0.99 (0.96, 1.00)	0.99**
SDANN(ms)	0.90 (0.73, 0.97)	0.86**	0.90 (0.71, 0.96)	0.82**	0.98 (0.95, 0.99)	0.97**
SDNNi (ms)	0.93 (0.80, 0.98)	0.94**	0.98 (0.95, 0.99)	0.97**	0.98 (0.94, 0.99)	0.99**
rMSSD (ms)	0.62 (−0.07, 0.86)	0.87**	0.83 (0.52, 0.94)	0.94**	0.96 (0.88, 0.99)	0.97**
pNN50 (%)	0.96 (0.89, 0.99)	0.98**	0.97 (0.91, 0.99)	0.98**	1.00 (0.99, 1.00)	1.00**
Frequency-domain						
VLF (ms^2)	0.77 (0.34, 0.92)	0.95**	−0.49 (−3.15, 0.48)	0.76**	0.29 (−0.99, 0.75)	0.79**
LF (ms^2)	0.82 (0.50, 0.94)	0.90**	0.91 (0.74, 0.97)	0.96**	0.98 (0.93, 0.99)	0.97**
HF (ms^2)	0.87 (0.64, 0.95)	0.96**	0.95 (0.87, 0.98)	0.97**	0.92 (0.79, 0.97)	0.95**
LF/HF ratio	0.93 (0.79, 0.97)	0.90**	0.96 (0.90, 0.99)	0.95**	0.89 (0.71, 0.96)	0.94**

Note. ICC = Intraclass Correlation Coefficient; r = Pearson Correlation Coefficient; CI = Confidence Interval; Mean RR = Mean beat-to-beat intervals; SDNN = Standard deviation of all RR intervals; SDANN = Standard deviation of the averages of RR intervals in all 5 min segments of the entire recording; SDNNi = Mean of the standard deviations of all RR intervals for all 5 min segments of the entire recording; rMSSD = Square root of the mean of the squares of differences between adjacent RR intervals; pNN50 = Proportion derived by dividing the number of interval differences of successive RR intervals greater than 50 ms by the total number of RR intervals; VLF = Very Low Frequency; LF = Low Frequency; HF = High Frequency.

** $p < .01$.

Table 3

Means and standard deviations of heart rate variability parameters across software programs.

	MARS	MindWare	Kubios	
	M (SD)	*M (SD)*	*M (SD)*	*F*
Mean RR (ms)	700.56 (51.88)	705.32 (48.57)	700.03 (48.87)	0.06
Time-domain				
SDNN (ms)	84.38 (21.81)	93.45 (23.59)	89.13 (22.01)	0.65
SDANN (ms)	40.50 (12.36)	43.69 (15.52)	41.94 (15.05)	0.20
SDNNi (ms)	73.00 (20.87)	81.51 (22.18)	76.41 (20.19)	0.66
rMSSD (ms)	50.56 (14.65)	70.08 (25.86)	62.18 (22.96)	3.28
pNN50 (%)	27.89 (12.79)	32.29 (14.20)	31.72 (13.91)	0.49
Frequency-domain				
VLF (ms^2)	1220.98 (586.32)	1822.60 (936.91)	3715.50 (1880.56)	17.10 ^{**abc}
LF (ms^2)	1280.83 (873.85)	1934.72 (1069.67)	1764.50 (927.43)	2.00
HF (ms^2)	996.22 (773.37)	1524.21 (1082.78)	1267.75 (737.91)	1.45
LF/HF ratio	1.49 (0.66)	1.70 (0.68)	1.35 (0.60)	1.15

Note. M = Mean; SD = Standard Deviation; F = F test-statistic from omnibus ANOVA; Superscript denotes follow-up pairwise comparison: ^{a}MARS vs. MindWare. ^{b}MARS vs. Kubios. ^{c}MindWare vs. Kubios; Mean RR = Mean beat-to-beat intervals; SDNN = Standard deviation of all RR intervals; SDANN = Standard deviation of the averages of RR intervals in all 5 min segments of the entire recording; SDNNi = Mean of the standard deviations of all RR intervals for all 5 min segments of the entire recording; rMSSD = Square root of the mean of the squares of differences between adjacent RR intervals; pNN50 = Proportion derived by dividing the number of interval differences of successive RR intervals greater than 50 ms by the total number of RR intervals; VLF = Very Low Frequency; LF = Low Frequency; HF = High Frequency.

** $p < .01$.