*Psychophysiology, 44* (2007), 203–215. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2006.00490.x

# Comparison of time and frequency domain measures of RSA in ambulatory recordings

ANNEBET D. GOEDHART,[a] SOPHIE VAN DER SLUIS,[a] JAN H. HOUTVEEN,[b] GONNEKE WILLEMSEN,[a] AND ECO J.C. DE GEUS[a]

[a]Department of Biological Psychology, Vrije Universiteit Amsterdam, The Netherlands
[b]Departement of Health Psychology, University of Utrecht, Utrecht, The Netherlands

**Abstract**

The extent to which various measures of ambulatory respiratory sinus arrhythmia (RSA) capture the same information across conditions in different subjects remains unclear. In this study the root mean square of successive differences (RMSSD), peak valley RSA (pvRSA), and high frequency power (HF power) were assessed during ambulatory recording in 84 subjects, of which 64 were retested after about 3 years. We used covariance structure modeling to test the equality of the correlations among three RSA measures over two test days and three conditions (daytime sitting or walking and nighttime sleep) and in groups with low, medium, and high mean heart rate (HR), or low, medium, and high mean respiration rate (RR). Results showed that ambulatory RMSSD, pvRSA, and HF power are highly correlated and that their correlation is stable across time, ambulatory conditions, and a wide range of resting HR and RR values. RMSSD appears to be the most cost-efficient measure of RSA.

**Descriptors:** Heart rate variability, Ambulatory, Parasympathetic, Interbeat interval, Respiration

Measures of heart rate variability (HRV) provide a window on the modulation of heart rate by the autonomic nervous system and have broad applications in both human and animal physiology (Berntson et al., 1997; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Within-subject studies show that HRV is responsive to changes in psychological state, particularly mental load and emotional stress (Allen & Crowell, 1989; Kamphuis & Frowein, 1985; Langewitz & Ruddel, 1989; Mulder, 1992; Sakakibara, Takeuchi, & Hayano, 1994), and to changes in posture and physical activity (Hatfield et al., 1998; Houtveen, Groot, & de Geus, 2005; Houtveen, Rietveld, & de Geus, 2002; Tulppo, Makikallio, Takala, Seppanen, & Huikuri, 1996). Between-subjects studies further show that lower levels of HRV independently predict cardiac disease and cardiac mortality (Bigger, Fleiss, Rolnitzky, & Steinman, 1993; Dekker et al., 1997, 2000; Hayano et al., 1991; Lombardi et al., 1987; Nolan et al., 1998; Saul et al., 1988; Singer et al., 1988; Singh et al., 1998; Tsuji et al., 1996.)

Most research has focused on HRV in the respiratory frequency range, also known as respiratory sinus arrhythmia (RSA). RSA is the difference in heart period during the inspiratory and expiratory phases of the respiratory cycle. RSA shows virtually no sensitivity to sympathetic nervous system activity but is affected in a dose–response way by muscarinergic blockers in humans (Martinmaki, Rusko, Kooistra, Kettunen, & Saalasti, 2006) or vagal cooling in animals (Katona & Jih, 1975). This has led to the use of tonic RSA levels as a proxy for individual differences in vagal cardiac control (Berntson et al., 1997; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996), although not without controversy because of potential confounding by individual differences in sensitivity of chemoreceptor and baroreceptor reflexes (Berntson et al., 1997; Houtveen et al., 2002) and by individual differences in respiratory behavior (Grossman & Kollai, 1993; Grossman, Wilhelm, & Spoerle, 2004; Ritz & Dahme, 2006).

RSA can be derived from the interbeat interval (IBI) time series in the time domain by taking the root mean square of differences between successive interbeat intervals (RMSSD; Penttila et al., 2001) or, in the frequency domain, by Fourier analysis (Akselrod et al., 1981, 1985) or Wavelet analysis (Pichot et al., 1999; Wiklund, Akay, & Niklasson, 1997). RSA can also be derived by peak–valley estimation (pvRSA; Katona & Jih, 1975) using the time series of IBIs in combination with the respiration signal.

An important feature of these time- and frequency-domain RSA measures is that they can be reliably measured under naturalistic conditions with the use of ambulatory monitoring (de Geus, Willemsen, Klaver, & van Doornen, 1995; Houtveen et al., 2005; Wilhelm, Roth, & Sackner, 2003). In stress research, ambulatory recording is a huge advantage. Different between-

Preparation of this article was, in part, financially supported by NWO/MaGW VIDI-016-065-318.

Address reprint requests to: Annebet D. Goedhart, Vrije Universiteit, Department of Biological Psychology, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. E-mail: ad.goedhart@psy.vu.nl.

subjects and within-subject mechanisms may determine cardiovascular reactivity to artificial laboratory stressors than to realistic stressors, encountered repeatedly at home or in the work setting. Generalization of individual differences in cardiovascular stress reactivity from standardized laboratory situations to actual real-life situations has indeed been shown to be moderate at best (Gerin, Rosofsky, Pieper, & Pickering, 1994; Kamarck, Schwartz, Janicki, Shiffman, & Raynor, 2003; Van Doornen, Knol, Willemsen, & de Geus, 1994).

With regard to potential negative health consequences of stress, ambulatory monitoring may be expected to have higher predictive validity for long-term health outcomes than laboratory measurements. This has already been shown to be the case for blood pressure, where ambulatory levels are better predictors for cardiovascular morbidity and mortality than laboratory or office measurements (Pickering & Devereux, 1987; Verdecchia et al., 1994, 1998; Verdecchia, Schillaci, Reboldi, Franklin, & Porcellati, 2001). It is reasonable to assume that prolonged recording of RSA in naturalistic settings will also have added predictive power over short-term recordings. This assumption, however, remains to be tested empirically.

Testing this assumption will require large-scale ambulatory recording in many thousands of subjects and a rational choice between the various RSA measures is direly needed. Unfortunately, the extent to which these measures capture the same information across different ambulatory conditions and different subjects remains unclear. Although time and frequency domain measures are highly correlated under standardized recordings, with $rs > .80$ (Bigger, Fleiss, Steinman, et al., 1992; Byrne & Porges, 1993; Grossman, van Beek, & Wientjes, 1990; Hayano et al., 1991; Houtveen & Molenaar, 2001; Penttila et al., 2001), we may expect them to diverge more strongly under ambulatory recording conditions. In ambulatory recordings, higher ecological validity is balanced by a lack of experimental control over important confounders of RSA. In comparison to laboratory recording, ambulatory settings are characterized by frequent changes in activity and posture, frequent speech, circadian rhythms, temperature variations, and a larger variance in emotional state and mental load. The differential sensitivity of the various RSA measures to these within-subject factors is currently unknown.

Previous studies have shown that the sharpest changes in RSA levels arise when going from awake to sleep recording, and that RSA in awake recordings is most sensitive to changes in physical activity and posture (Grossman et al., 2004; Kupper et al., 2004; Vrijkotte, van Doornen, & de Geus, 2000). One design to examine the potential differential sensitivity of various RSA measures to these within-subject factors is to compare the structure of the correlations between ambulatory RSA measures across daytime and nighttime recordings, and, during the daytime part of the recording, across sitting activities and activities involving upright physical activity.

Even within these restricted ambulatory conditions, between-subject variance in the mean and range of respiration rate (RR) and heart rate may still be larger than in laboratory testing. This is problematic because individual differences in RR and heart rate both influence RSA measures, and such influence may be independent of cardiac vagal control (Berntson, Lozano, & Chen, 2005; Grossman, Karemaker, & Wieling, 1991; Grossman & Kollai, 1993). The differential sensitivity of the various RSA measures to these between-subject confounders is currently unknown. It can be addressed by comparing the correlation structure of ambulatory RSA measures across groups of subjects selected to have low, medium, and high mean heart rate during the ambulatory test day, or across groups with low, medium, and high mean RR.

In the present study we test the correlation between RMSSD, peak–valley RSA (pvRSA), and a frequency domain (HF power) measure of RSA in an ambulatory setting. First, reliability will be assessed for each of the measures by looking at short-term within-day test–retest correlations and correlations between genetically identical twins. Next, temporal stability will be assessed across an average period of 3 years. Finally, we will use covariance structure modeling to test the equality of the correlations across major changes in ambulatory activity (sleep vs. awake, sitting vs. awake standing/walking) and across groups of subjects with low, medium, and high mean IBI or low, medium, and high mean RR. Based on previous laboratory findings and a previous small-scale ambulatory study (Vrijkotte, Riese, & de Geus, 2001), we expect that RMSSD, pvRSA, and HF power will show high correlations over time, across ambulatory activity, and across a wide range of mean heart rate and RR.

## Methods

### *Participants*

Participants were all registered in the Netherlands Twin Register (NTR). They came from families that participated in a linkage study searching for genes influencing personality and cardiovascular disease risk, which is described elsewhere (Boomsma et al., 2000). Out of the 1332 twins and siblings who returned a DNA sample (buccal swabs) for the linkage study, 816 also participated in cardiovascular ambulatory monitoring. Reasons for exclusion were pregnancy, heart transplantation, pacemaker and known ischemic heart disease, congestive heart failure, or diabetic neuropathy. Of these participants a total of 65 (20 male, 45 female) participants were tested twice, separated by a minimum of 2 years and 1 month and a maximum of 4 years and 8 months (mean 3 years and 4 months). RSA measures could not be reliably obtained in 1 participant. Age at the first day of testing in the remaining 64 participants ranged between 18 and 62 years (mean = 30.9, $SD = 9.7$). In addition, 20 randomly selected identical MZ twin pairs (18 male, 22 female) with zygosity confirmed by DNA typing, were also included. Average age of the twins was 27 with a range of 18 to 32. These twins were only tested once. The Ethics Committee of the Vrije Universiteit approved the study protocol and all participants gave written consent before entering the study. No payment was made for participation, but all subjects received an annotated review of their ambulatory heart rate recordings.

### *Ambulatory Recording*

Subjects were invited to participate in the study by letter and all participants were subsequently phoned by the researchers, who provided additional information on the study and made an appointment with the participants for 24-h ambulatory monitoring. The first ambulatory measurement took place during a representative workday (or a day with representative housekeeping chores for those who were not employed). The second ambulatory measurement day took place during a comparable (work) day for most of the participants, but 17 subjects would only participate if the repeated measurement was scheduled on a leisure day. On the day preceding monitoring and on the

monitoring day itself participants were asked to refrain from leisure time exercise or heavy physical work. Participants were visited at home between 7:00 and 10:00 a.m. and were fitted with the Vrije Universiteit Ambulatory Monitoring System (VU-AMS46; de Geus et al., 1995; Riese et al., 2003; Willemsen, de Geus, Klaver, van Doornen, & Carroll, 1996). The VU-AMS produced an audible alarm approximately every 30 min (10 min randomized) to prompt the participant to fill out an activity diary. Participants were instructed to write down their physical activity and bodily postures during the last 30-min period in chronological order. Diary prompting was disabled during sleep.

The ECG and changes in the thorax impedance (dZ) were recorded continuously using six disposable, pregelled Ag/AgCl electrodes. The first ECG/dZ electrode was placed on the sternum over the first rib between the two collarbones. The second ECG electrode was placed at the apex of the heart over the ninth rib on the left lateral margin of the chest approximately 3 cm under the left nipple. The third ECG electrode is a ground electrode and was placed at the lower right abdomen. A second dZ measuring electrode was placed over the tip of the xiphoid complex of the sternum. The dZ current electrodes were placed on the back over cervical vertebra C4 and between thorax vertebras T8 and T9. Electrode resistance was kept low (below 10 kΩ) by cleaning the skin with alcohol and rubbing.

### *Ambulatory Signal Scoring*

Using the activity diary entries in combination with a visual display of the output of an inbuilt vertical accelerometer, the entire 24-h recording was divided into fixed periods. These periods were coded for posture (supine, sitting, standing, walking, bicycling), physical activity (e.g., desk work, dinner, meetings, watching TV), and physical load (no load, light, intermediate, and heavy). Minimum duration of periods was always 5 min and maximum duration was always 1 h. If periods with similar activity and posture lasted more than 1 h (e.g., during sleep), they were divided into multiple periods of maximally 1 h. All periods were classified as lying asleep, sitting, or standing/walking based on the dominant posture reported; the exact timing of changes in posture was verified using the accelerometer signal from the ambulatory device. We then looked at the self-reported activity and physical load to determine whether this period could be classified as sitting with light physical activity (desk work, watching TV, writing, eating, reading, etc.) or sitting interspersed with intermediate physical activity (machine operation). The periods interspersed with intermediate activity were discarded. For standing/walking periods we selected only those periods in which the participants reported no more than light physical load. For each period coded for posture, activity, and physical load we determined the average RMSSD, pvRSA, and HF power. An average of 25 periods was created per participant. The mean duration of the sleep periods was 56 min ($SD = 11$), of the sitting activities 26 min ($SD = 14$), and of the standing/walking condition 19 min ($SD = 12$). This procedure allowed us to test the sensitivity of the correlation structure of the three RSA measures to major changes in posture and activity.

From the ECG and the dZ we obtained the IBI time series and respiration signal according to the procedures detailed elsewhere (de Geus et al., 1995). Artifact preprocessing was performed on the IBI data. When the IBI deviated more than 3 *SD* from the moving mean of a particular period, it was automatically identified as an artifact and accepted or overruled by visual inspection. Because artifacts cannot simply be deleted because the continuity of time would be lost, spuriously short IBIs were summed and missing beats were "created" by splitting spuriously long IBIs. The mean IBI and RMSSD values were computed from these corrected IBI time series across each of the labelled periods. RMSSD was defined as

$$RMSSD = \sqrt{\frac{1}{n}\sum (IBI_i - IBI_{i-1})^2}.$$

Per breath, estimates of pvRSA were obtained by substracting the shortest IBI during heart rate acceleration in the inspirational phase (which was made to include 750 ms from the following expiration to account for phase shifts) from the longest IBI during deceleration in the expirational phase (including 750 ms from the following expiratory pause/inspirational phase). When no phase-related acceleration or deceleration was found, the breath was assigned a pvRSA score of zero. Automatic scoring of RR and pvRSA was checked by visual inspection of the respiratory signal from the entire recording. Breathing cycles that showed irregularities like gasps, breath holding, or coughing were not considered valid and were rejected and removed from further processing. The 3% shortest and longest breaths were automatically removed from the entire recording before averaging pvRSA across all remaining breaths to a single mean pvRSA for each of the labeled periods. We discarded 17.3% of all automatically scored breaths. A total of 10.3% of these breaths occurred in periods in which we could not reliably establish posture and activity or in which signal quality was deemed insufficient during visual inspection. A further 7% were nonplausible long or short breaths, deviated more than 3 *SD* from the mean, or had close to zero amplitude.

Computation of HF power by Fourier analysis, a widely used strategy to asses RSA, assumes that the data show at least weak stationarity (Weber, Molenaar, & van der Molen, 1992). Stationarity of time series may be interpreted as having a stable mean and variance over time. In ambulatory studies and/or for analysis of relatively long data segments, the assumption of stationarity is likely to be violated. Therefore, we improve on the usual Fourier approach by using a Wavelet decomposition for the computation of HF power that does not have a stationarity assumption (Houtveen & Molenaar, 2001). Additionally, by using Wavelet transformation, much longer ambulatory fragments can be selected for cross-method comparison. Uniformly spaced samples were created by interpolation of the IBI data using a Wavelet interpolation algorithm. Next, Discrete Wavelet Transformation (DWT) was performed using a cardinal cubic spline function as base (see Houtveen & Molenaar, 2001, for more information regarding this procedure). This method results in identical power values for stationary relatively short coded periods (e.g., 7 min of quiet reading) as compared to Fourier transformation, but it is superior for our relatively longer and nonstationary coded periods (e.g., first hour of sleep). The HF power was computed as the sum of the variances of the 0.125–0.25-Hz and 0.25–0.5-Hz windows. Note that the size of a frequency window always doubles after each Wavelet decomposition step. Because the DWT (like Fourier) suffers from aliasing effects at both ends, the first and last 40 data points (2.5 s) of the time series were excluded from the derivation of the variances.

### *Statistical Analyses*

*Reliability, heritability and temporal stability*. To test the reliability of RMSSD, pvRSA, and HF power, the short-term

within-day test–retest correlations and the correlations between genetically identical twins were assessed. The within-day correlations were computed between the second and the third hour of sleep and also between two periods of comparable sitting activities during daytime recordings (e.g., reading a magazine or newspaper or watching TV).

The MZ correlations and temporal stabilities were separately computed for three main ambulatory conditions (sleep, awake sitting activities, awake standing/walking). When the MZ correlation is not unity, this means that environmental influences are creating dissimilarity between genetically identical subjects. Measurement error is completely contained within these environmental influences. Hence, the MZ correlation sets an upper limit to measurement error corresponding to $[1rMZ]^2$. Temporal stability was computed as the intraclass correlation between the first measurement and the second measurement that took place after an average period of 3 years and 4 months.

*Correlations between RMSSD, pvRSA, and HF power.* We used covariance structure modeling (Bollen, 1989) to test four hypotheses regarding the equality of the correlations among the three measures (RMSSD, pvRSA, and HF power). All modeling was performed in Mplus, version 4 (Muthén & Muthén, 2005). LISREL 8.53 (Jöreskog & Sörbom, 2001) was used to calculate the standardized residuals.

The first hypothesis states that the correlation structure remains stable over time, that is, across the two test days. To test this hypothesis, we first estimated the full correlation matrix between RMSSD, pvRSA, and HF power in all ambulatory conditions on Test Day 1 and Test Day 2. This model, in which all relations between the measures are estimated freely, is saturated in the sense that the number of estimated parameters equals the number of observed statistics. The saturated model therefore fits the data perfectly, that is, had zero degrees of freedom, and a $\chi^2$ of exactly zero. Subsequently, the correlations between the three measures (RMSSD, pvRSA, and HF power) collected during sleeping, sitting, and standing/walking on Day 1 were constrained to be equal to the symmetric correlations collected on Day 2. In addition, the cross-measure cross-test day correlations were fixed to be equal (i.e., the correlations of the measures collected on Test Day 1 with those collected on Test Day 2 were fixed to be equal to the correlations of the measures collected on Test Day 2 with those collected on Test Day 1). Figure 1 illustrates this testing procedure.

Under the saturated baseline model, illustrated in the upper panel of Figure 1, all 153 correlations are estimated freely, as is denoted by all correlations having different indices (i.e., all correlations between measures collected on Test Day 1 have index A, all correlations between measures collected on Test Day 2 have index B, all test–retest correlations have index X, and the cross-measure cross-test day correlations have either index C or D). Under the alternative, more restricted, model illustrated in the lower panel of Figure 1, the 36 correlations between the measures collected on Test Day 1 are constrained to be equal to the 36 correlations between the measures collected on Test Day 2. In the lower panel of Figure 1, these correlations all have index A, whereas correlations with the same numerical extension are actually constrained to be equal (i.e., correlation A1 on Test Day 1 is set equal to correlation A1 on Test Day 2, etc.). In addition, the cross-measure cross-test day correlations are set to be equal. Therefore, in the lower panel of Figure 1, the 36 correlations of the measures collected on Test Day 1 with the measures collected on Test Day 2 have the same index C as the 36 correlations of the measures collected on Test Day 2 with the measures collected on Test Day 1. Again, correlations with the same numerical extension are fixed to be equal. All in all, this resulted in an alternative model with 72 constraints, and thus 72 degrees of freedom.

The second hypothesis states that the correlation structure is stable over the three main ambulatory conditions (sleep, awake sitting, awake walking). Here, ambulatory data were available for 84 subjects, the 64 subjects that were tested twice and an additional 20 subjects obtained by randomly selecting one of the twins from the 20 MZ twins pairs that were used to compute the MZ twin correlations. The testing procedure with respect to the effect of ambulatory condition is illustrated in Figure 2. Under the saturated baseline model, illustrated in the upper panel of Figure 2, all 36 correlations are estimated freely, as is denoted by all correlations having different indices. More specifically, all correlations between measures collected in the same ambulatory condition have either index A (sleeping), B (sitting), or C (standing/walking). All test–retest correlations of the same measures collected across different ambulatory conditions have index X (between sleeping and sitting), Y (between sleeping and standing/walking), or Z (between sitting and standing/walking). All cross-measure cross-condition correlations have indices D, E, F, G, H, and I, respectively. Under the alternative, more restricted model illustrated in the lower panel of Figure 2, the correlations between the three measures (RMSSD, pvRSA, and HF power) are constrained across ambulatory conditions (index A), such that correlations with similar numerical extensions are equal (e.g., all A1s are equal). In addition, the cross-measure cross-condition correlations are constrained (index D), such that correlations with the same numerical extension are equal (e.g., all D1s are equal). This resulted in an alternative, restricted model with 21 constraints, and thus 21 degrees of freedom.

The third and fourth hypotheses state that the correlation structure is independent of mean IBI and RR, respectively. These hypotheses were tested on the data obtained in the 84 subjects on the first test day. We subdivided this sample first into three IBI and next into three RR groups. To do so, mean IBI and RR scores were calculated for each participant across the three ambulatory conditions. Based on these mean scores, we distinguished "low," "medium," and "high" IBI and RR groups, corresponding to the lowest 33%, the medium 33%, and the highest 33% of the sample. To test whether the correlation matrices were equal for the low, medium, and high groups, multigroup analyses were conducted, in which the correlations between the nine measures recorded on Test Day 1 (three measures collected under three ambulatory conditions) were first estimated freely in all groups (saturated model), and then constrained to be equal across the groups. For instance, in the alternative model for the IBI groups, the 36 correlations of the low IBI group were constrained to be equal to the 36 correlations of the medium IBI group, and the 36 correlations of the high IBI group, resulting in a restricted model with 72 degrees of freedom. Note that the restrictions concerned the correlations and not the covariances, so the variances of the measures were allowed to differ across groups.

*Fit statistics.* In testing Hypotheses 1 to 4, the more restricted models are always nested under the saturated model (Bollen, 1989). Normally, the fit of nested models is evaluated by means of a likelihood ratio test. This test is constructed by subtracting the $\chi^2$ value of the less restrained model with more freely

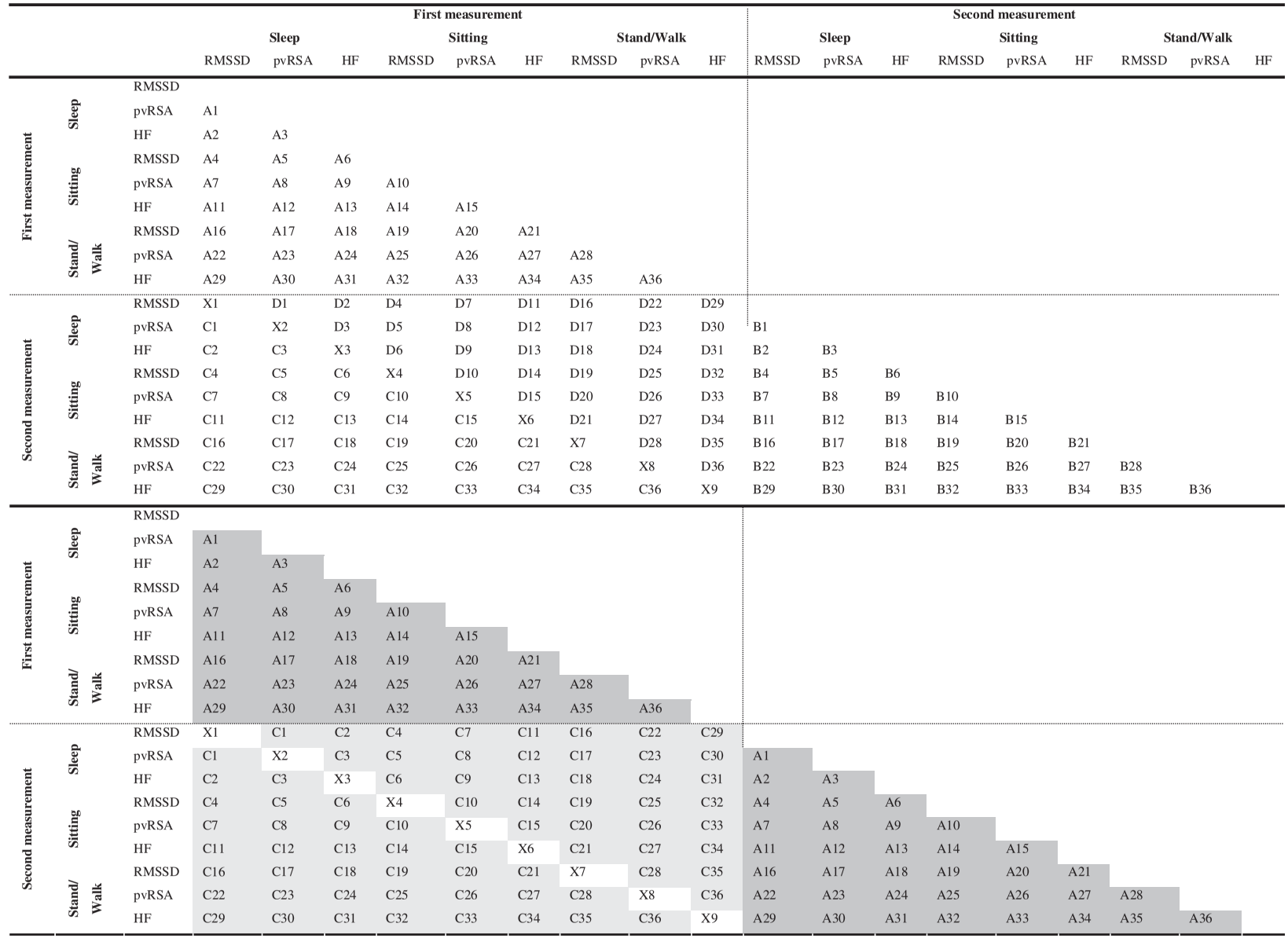


**Figure 1.** Testing equality of the correlations of the RSA measures across the two test days. The upper panel shows the Null model in which all 153 correlations are estimated freely. The lower panel shows the Alternative model in which correlations with the same indices are constrained to be equal.

| | | | First measurement | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Sleep | | | Sitting | | | Stand/Walk | | |
| | | | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF |
| First measurement | Sleep | RMSSD | | | | | | | | | |
| | | pvRSA | A1 | | | | | | | | |
| | | HF | A2 | A3 | | | | | | | |
| | Sitting | RMSSD | X1 | E1 | E2 | | | | | | |
| | | pvRSA | D1 | X2 | E3 | B1 | | | | | |
| | | HF | D2 | D3 | X3 | B2 | B3 | | | | |
| | Stand/Walk | RMSSD | Y1 | G1 | G2 | Z1 | I1 | I2 | | | |
| | | pvRSA | F1 | Y2 | G3 | H1 | Z2 | I3 | C1 | | |
| | | HF | F2 | F3 | Y3 | H2 | H3 | Z3 | C2 | C3 | |
| First measurement | Sleep | RMSSD | | | | | | | | | |
| | | pvRSA | A1 | | | | | | | | |
| | | HF | A2 | A3 | | | | | | | |
| | Sitting | RMSSD | X1 | D1 | D2 | | | | | | |
| | | pvRSA | D1 | X2 | D3 | A1 | | | | | |
| | | HF | D2 | D3 | X3 | A2 | A3 | | | | |
| | Stand/Walk | RMSSD | Y1 | D1 | D2 | Z1 | D1 | D2 | | | |
| | | pvRSA | D1 | Y2 | D3 | D1 | Z2 | D3 | A1 | | |
| | | HF | D2 | D3 | Y3 | D2 | D3 | Z3 | A2 | A3 | |

**Figure 2.** Testing equality of the correlations of the RSA measures across ambulatory conditions. The upper panel shows the Null model in which all 36 correlations are estimated freely. The lower panel shows the Alternative model in which correlations with the same indices are constrained to be equal.

estimated parameters, from the $\chi^2$ value of the more restricted model with fewer free parameters. The difference in $\chi^2$ (denoted as $\chi^2_{\text{diff}}$) between the two models follows a $\chi^2$ distribution with the number of degrees of freedom (*df*) equaling the difference in the number of parameters estimated in the two models. The restricted model is considered tenable if its value of the $\chi^2$ goodness-of-fit statistic is not significantly greater than that of the more lenient model, that is, if the difference in $\chi^2$ is not significant. However, Browne, MacCallum, Kim, Andersen, and Glaser (2002) noted that the $\chi^2$ statistic can be markedly inflated if some measures in the model are highly correlated, for example, if highly reliable measures are used to measure the same or related characteristics several times. In that case, standardized residuals, which are a function of the differences between the observed covariance matrix **S** and the estimated covariance matrix $\Sigma$, may be (very) small, indicating close fit of the model to the observed data, whereas the $\chi^2$ statistic, and fit indices based on this statistic, indicate a poorly fitting model.

We expected that many of the correlations between the dependent measures in this study would be larger than .80. Although such high correlations are desirable in the sense that they indicate reliable measurement and substantial overlap between measures, the consequence may be that the $\chi^2$ statistic of the restricted models may assume large values in the presence of trivial misfit. Comparing the restricted model to the more lenient model using the usual likelihood ratio test may then result in the rejection of a perfectly acceptable model. In view of the above, we chose to evaluate the fit of the restricted model in a different way. If the fit of the restricted model was in itself acceptable, we always considered the constraints to be tenable, that is, the sets of correlations to be equal. Although we will report the $\chi^2$ statistic of every tested model to conform to common practice, our main indices of fit were the Comparative Fit Index (CFI) and the standardized residuals (Bollen & Long, 1993; Schermelleh-Engel, Moosbrugger, & Müller, 2003). The CFI is based on the $\chi^2$ statistic, but introduces penalties for every additional parameter estimated (i.e., favors more parsimonious models) and is relatively unaffected by sample size. The CFI ranges between 0 and 1.00, with values below .95 indicating poor fit, values between .95 and .97 indicating acceptable fit, and values between .97 and 1.00 indicating good fit. Standardized residuals are standardized differences between the observed covariance matrix **S** and the estimated covariance matrix $\Sigma$. Ideally, the standardized residuals should lie between $-3$ and $+3$ and show a normal distribution by approximation (this can be evaluated readily using LISREL's stem leaf plots). In case the CFI is below .95 and/or the standardized residuals are outside the acceptable range ($-3$ and $+3$), the largest (absolute) standardized residual usually indicates the element that is most poorly fitted by the model. To trace these sources of local misspecification, we planned to use the Modification Indices (MIs) supplied by the Mplus program. MIs are calculated for every fixed parameter in the model, and the value of the MIs represents the expected drop in overall $\chi^2$ (i.e., improvement in model fit) if the parameter were to be freely estimated.

*Missingness*. In the comparison across ambulatory conditions, complete data during sleep on both days were available for

**Table 1.** *Means (SD) of RMSSD, pvRSA, and HF Power Separately for the Two Test Days*

| | *N* | RMSSD (ms) | pvRSA (ms) | HF ($ms^2$) | IBI (ms) | RR (bpm) |
|---|---|---|---|---|---|---|
| Full recording | | | | | | |
| Test | 84 | 47.60 (24.06) | 49.82 (22.08) | 786.22 (705.55) | 803.99 (10.44) | 16.70 (1.10) |
| Retest | 64 | 41.71 (20.85) | 43.95 (19.87) | 618.68 (562.17) | 792.80 (83.65) | 16.63 (1.35) |
| Sleep | | | | | | |
| Test | 78 | 65.65 (33.19) | 60.86 (31.22) | 1109.05 (1097.01) | 981.49 (126.35) | 16.23 (2.00) |
| Retest | 57 | 60.36 (36.95) | 51.40 (25.49) | 991.11 (1163.15) | 970.61 (107.32) | 16.17 (2.15) |
| Sitting | | | | | | |
| Test | 84 | 42.00 (24.34) | 50.10 (23.97) | 688.50 (686.82) | 772.67 (109.60) | 17.00 (1.25) |
| Retest | 64 | 39.47 (22.67) | 46.59 (24.41) | 590.30 (622.00) | 791.15 (87.17) | 16.84 (1.70) |
| Standing/walking | | | | | | |
| Test | 84 | 38.04 (920.52) | 40.35 (19.44) | 590.63 (545.34) | 689.31 (97.25) | 16.87 (1.01) |
| Retest | 64 | 33.43 (14.73) | 36.77 (15.64) | 444.61 (344.60) | 700.33 (70.57) | 16.72 (1.25) |

51 subjects only, due to ECG or ICG signal loss during sleep. In the presence of missing data, one can use Full Information Maximum Likelihood (FIML) estimation, which uses all available data. However, this method of estimation should only be applied if data are missing (completely) at random (MAR or MCAR; we refer the reader to Schafer & Graham, 2002, for a detailed discussion of mechanisms of missingness). The missingness in our data was therefore first examined with SPSS missing data analysis. When considered across all 18 measures, missingness could be considered completely at random (MCAR), as indicated by the nonsignificance of Little's MCAR test ($\chi^2(80) = 99.83$, n.s.). In both Mplus and LISREL, FIML estimation could therefore be used to accommodate missing data so that all available data were used in the model estimation.

## Results

### *Means*

Table 1 presents the untransformed means and standard deviations for RMSSD, pvRSA, HF power, IBI, and RR across all ambulatory conditions and separately for each ambulatory condition (i.e., sleep, awake sitting, and awake standing/walking). Because the RMSSD, pvRSA, and HF power distributions were skewed, their natural logarithms were used in all further analyses. Repeated measures ANOVA showed a significant effect of ambulatory condition on RMSSD, $F(2,51) = 31.63$, $p < .001$, pvRSA, $F(2,50) = 35.94$, $p < .001$, and HF power, $F(2,51) = 10.48$, $p < .001$, and post hoc testing showed that all three measures decreased significantly from sleep to sitting to standing/walking, all $p < .05$. There were no significant main effects of test day on the means of RMSSD, pvRSA, and HF power or interaction effects between ambulatory condition and test day.

### *Short-Term Reliability*

Within-day correlations between the second and the third hour of sleep and between two periods of sitting activities all exceeded .85 (see Table 2) suggesting good to excellent short-term reliability for RMSSD, pvRSA, and HF power alike. For comparison, short-term reliability of IBI and RR is also given.

The intrapair MZ correlations further confirmed good reliability (see Table 3). For all three measures, MZ correlations were highest during sitting activities and lowest during standing/walking. Even at standing/walking, however, the lowest MZ correlation for pvRSA suggests that measurement error cannot account for more than 18% of the variance ($[1 - .58]^2$). Based on the within-day test–retest or MZ twin correlations, none of the three measures could be favored as the "best," that is, most reliable, RSA measure.

### *Temporal Stability*

Table 4 displays the correlations across the two test days for the three RSA measures, IBI and RR. Temporal stability for RMSSD, pvRSA, and HF power was good when computed across all ambulatory conditions and separately across awake sitting activities. Temporal stability was moderate during standing/walking, potentially because of the low stability of the respiratory frequency during these periods of physical activity. For all measures, recordings proved most stable during sleep. As with short-term reliability, none of the three measures seemed to be clearly favored by temporal stability as the best RSA measure.

### *Correlations between RMSSD, pvRSA, and HF Power*

Table 5 presents the full correlation matrix between the three RSA measures in all ambulatory conditions on Test Day 1 (upper left block) and Test Day 2 (lower right block) and across test days (lower left block). As can be seen in Table 5, many of the correlations between the dependent measures in this study were larger than .80, and quite a few exceed .90. This justifies our approach of evaluating the fit of the restricted models directly besides comparing the fit of the restricted models to the fit of the saturated model.

*Effect of test day*. The first hypothesis states that the correlation structure remains stable over time, that is, across the two test days. The fit indices of the alternative model representing this hypothesis are shown in Table 6 (Model TD). As expected, the difference in $\chi^2$ between the saturated model and the alternative model was significant, $\chi^2_{diff}(72) = 130.12$, $p < .001$. Yet, both the CFI and the standardized residuals indicated that the fit of the alternative model was good. We therefore conclude that the correlations can, to reasonable approximation, be considered equal across the two test days.

*Effect of ambulatory conditions*. The second hypothesis states that the correlation structure is stable over the three main ambulatory conditions (sleep, awake sitting, awake standing/walking). Taken that the effect of test day was negligible, the effect of ambulatory condition on the correlations among RMSSD, pvRSA, and HF power was initially tested on data from the first test day only (Model AMBa in Table 6). As expected, the difference in $\chi^2$ between the restricted model and the saturated model was significant, $\chi^2(21) = 95.53$, $p < .001$. However, the

**Table 2.** *Within-Day Correlations (Intraclass Correlations) between the Second and the Third Hour of Sleep and between Two Periods of Light Physical Sitting Activities*

| | *N* | RMSSD (ms) | pvRSA (ms) | HF ($ms^2$) | IBI (ms) | RR (bpm) |
|---|---|---|---|---|---|---|
| Sleep | 150 | 0.88** | 0.86** | 0.89** | 0.92** | 0.93** |
| Sitting | 155 | 0.86** | 0.85** | 0.87** | 0.84** | 0.64** |

**Correlation is significant at .01 level.

CFI was only just below the critical value of .95 (CFI = .94), and the standardized residuals were clearly within the acceptable range. We repeated the analysis on the data from the second test day (Model AMBb). Again, the difference in $\chi^2$ between the restricted model and the saturated model was significant, $\chi^2(21) = 73.35$, $p < .001$, but the small standardized residuals and high CFI indicated good fit. Taken together, the analyses across the two days suggest invariance of the correlation structure of RMSSD, pvRSA, and HF power correlations across sleep, sitting, and standing/walking activities.

*Effect of mean IBI.* Next we tested whether the correlations among RMSSD, pvRSA, and HF power were equal for subjects, with low (707.09 ± 43.90), medium (795.95 ± 21.37), or high (922.71 ± 90.52) mean ambulatory IBI. These three groups differed significantly in their IBI scores, $F(2,81) = 92.77$, $p < .001$. Given that the effect of test day was negligible, the effect of group membership on the correlations among RMSSD, pvRSA, and HF power was tested on the data collected on the first test day only. The 36 correlations within each group were constrained to be equal for the low, medium, and high IBI groups, yielding a restricted model with 72 degrees of freedom (Model IBI). The fit indices of this restricted model are shown in Table 6. Although the difference in $\chi^2$ between the saturated model and the alternative was significant, $\chi^2_{diff}(72) = 107.54$, $p < .001$, the CFI and the standardized residuals indicated that the fit of the alternative model was acceptable. We therefore conclude that the correlations between the RMSSD, pvRSA, and HF power measures of RSA can be considered approximately equal across groups distinguished with respect to their mean ambulatory heart rate.

*Effect of mean RR.* Finally, we tested whether the correlations among RMSSD, pvRSA, and HF power were equal for subjects, with low (15.55 ± 0.45), medium (16.72 ± 0.25), or high (17.88 ± 0.68) mean ambulatory RR. These three groups differed significantly with respect to their RR scores, $F(2,81) = 155.09$, $p < .001$. Again, the effect of group membership on the correlations among RMSSD, pvRSA, and HF power was tested on the data collected on the first test day only as the effect of test day had proven negligible. The constraints imposed on the correlation matrices of the RR groups were analogous to those imposed on the matrices of the IBI groups. That is, the 36 correlations within each group were constrained to be equal for the low, medium, and high RR groups, yielding a restricted model with 72 degrees of freedom (Model RR). The fit indices of this restricted model are shown in Table 6. Although the model fitted significantly worse than the saturated model, $\chi^2_{diff}(72) = 136.67$, $p < .001$, both the CFI and the standardized residuals indicated good fit. We therefore conclude that the correlations between the RMSSD, pvRSA, and HF power measures can be considered approximately equal across groups distinguished with respect to their mean ambulatory RR.

## Discussion

Cardiovascular psychophysiology aimed at identifying individuals at risk for future cardiovascular disease is increasingly relying on ambulatory monitoring under the expectation that this has higher predictive validity for long-term health outcomes than laboratory measurements (Goldstein, Shapiro, & Guthrie, 2006; Grossman, 2004; Vrijkotte et al., 2000). RSA is a promising measure for large-scale ambulatory studies because it has been linked, both theoretically and empirically, to activity of the parasympathetic nervous system, that, in turn, is paramount to the electrical stability of the heart (Ando et al., 2005; Hull et al., 1990; Levy & Schwartz, 1994; Vanoli et al., 1991).

However, RSA can be assessed in many different ways, which clearly differ in terms of costs and whether they are cumbersome to the study participants or require labor-intensive data reduction by the researcher. A full ECG recording (e.g., the Holter monitor), for instance, is only mildly cumbersome to the patients, who need to wear electrodes and a small portable recording device on the hip. However, this mode of assessment is labor intensive to the researcher, who needs to perform repeated Fourier analyses to compute HF power on all stationary 5-min segments, which often need visual inspection of the automatically detected erroneous IBIs. In contrast, computation of the RMSSD from the IBI time series obtained from a wrist-watch type HR-recording device together with a single elastic recording band around the chest (e.g., the Polar Sporttester) is much less demanding of both participant and researcher. The researcher still needs to visually inspect the automatically detected erroneous IBIs, but computation of time domain measures like RMSSD is otherwise straightforward. Additional recording of the RR

**Table 3.** *Intrapair MZ Twin Correlations*

| | *N* | RMSSD (ms) | pvRSA (ms) | HF ($ms^2$) | IBI (ms) | RR (bpm) |
|---|---|---|---|---|---|---|
| Full recording | 20 | 0.76** | 0.75** | 0.76** | 0.74** | 0.61** |
| Sleep | 20 | 0.63** | 0.69** | 0.65** | 0.70** | 0.84** |
| Sitting | 20 | 0.81** | 0.80** | 0.80** | 0.82** | 0.42* |
| Standing/walking | 19 | 0.63** | 0.58** | 0.60** | 0.61** | 0.72** |

*Correlation is significant at .05 level.
**Correlation is significant at .01 level.

**Table 4.** *Temporal Stability for an Average Period of 3 Years and 4 Months*

| | *N* | RMSSD (ms) | pvRSA (ms) | HF ($ms^2$) | IBI (ms) | RR (bpm) |
|---|---|---|---|---|---|---|
| Full recording | 64 | 0.71** | 0.58** | 0.76** | 0.58** | 0.74** |
| Sleep | 52 | 0.73** | 0.72** | 0.81** | 0.65** | 0.91** |
| Sitting | 64 | 0.70** | 0.68** | 0.80** | 0.66** | 0.69** |
| Standing/walking | 64 | 0.44** | 0.44** | 0.57** | 0.61** | 0.37** |

**Correlation is significant at .01 level.

would allow computation of pvRSA, but at the cost of adding another layer of work to the data-reduction phase and an additional burden on the participant by requiring the wearing of either additional electrodes or respiratory bands.

In this study, we tested whether the assessment of between-subject differences in RSA was sensitive to the method used, that is, whether ''high tech'' pvRSA or HF power were superior to ''low tech'' RMSSD. The answer is a resounding *no*. All three RSA measures were highly correlated among each other and none of them stood out in terms of short-term reliability or temporal stability over a period of more than 3 years. The present findings with respect to the reliability of ambulatory RSA, either defined as a within-day short-term retest coefficient (.85–.89) or as the intrapair resemblance in genetically identical twins (.58–.81), are in line with previous studies that showed similarly high test–retest correlations for the average 24-h levels of RMSSD (.67–.89) and HF power (.76–.92) after 3 to 65 days in both healthy individuals and cardiac patients (Bigger, Fleiss, Rolnitzky, & Steinman, 1992; Hohnloser, Klingenheben, Zabel, Schroder, & Just, 1992; Kleiger et al., 1991; Sinnreich, Kark, Friedlander, Sapoznikov, & Luria, 1998; Stein, Rich, Rottman, & Kleiger, 1995). Likewise, the finding that RMSSD in the full recording is temporally stable (.71) across an average period of 3 years and 4 months is in agreement with the only previous study that had a prolonged test–retest interval and reported long-term stability of .79 for the 24-h RMSSD level (Pitzalis et al., 1996). The present results contribute to these previous studies by showing similar stability for HF power and pvRSA. In addition, the present study shows that the high intercorrelations of the three RSA measures do not change over prolonged periods of time. This suggests that longitudinal studies of RSA can, for instance, use pvRSA at the first wave and RMSSD at the second wave.

In keeping with a large body of literature, the mean values of the three RSA measures increased from standing to sitting and from sitting to sleep (Grossman et al., 2004; Houtveen et al., 2005; Martinmaki et al., 2006). Although temporal stability was good to excellent for sitting and sleep, it was only moderate for standing/walking. One possible explanation might be that it is more difficult to arrive at a reliable measure of RSA during standing/walking because the standing/walking periods are relatively short. The mean duration of the sleep periods was 56 min, whereas the mean duration was 26 min for sitting activities, and 19 min for standing/walking. However, if averaging RSA measures across longer periods would yield higher stability than averages across shorter periods, then we would expect the temporal stability of the RSA measures to be highest during sleep. This is not the pattern that is evident from Table 3, which shows that correlations during sleep and sitting are comparably high, even though the mean duration of sitting periods was only half the duration of the sleeping periods. Duration per se, therefore, does not seem to explain why RSA measures recorded during walking/standing are less stable than measures obtained in both other conditions. As an alternative explanation, the temporal stability of RSA during standing/walking activities may be lower than that of sitting and sleep, because respiratory behavior in this condition is much more variable. This is directly supported by the lower temporal stability of the respiratory frequency during these periods of physical activity. Based on extensive ambulatory pvRSA data, Grossman et al. (2004) have shown that physical activity needs to be taken into account when interpreting ambulatory RSA, and our data underscore their warning.

Independent of ambulatory condition, large differences in mean respiratory frequency and heart rate were found. We were concerned that such differences might distort the correlations between the three RSA measures, because RR and heart rate may both distort their relation to cardiac vagal control (Berntson et al., 2005; Grossman et al., 1991; Grossman & Kollai, 1993). This concern was greatly mitigated by the data. Differences in mean resting heart rate or mean RR did not affect the correlations between the various RSA measures; the RSA measures were as highly correlated in a group with a mean heart rate of 83 bpm as in a group with a mean heart rate of 65 bpm, and as highly correlated in groups with a mean RR of 15 (range 14–16) versus 18 (range 17–20) times per minute. This contradicts the idea that one of these RSA measures is relatively more sensitive than the others to confounders such as individual differences in RR (Grossman, 1992) or in heart rate (Berntson et al., 2005).

Taken together, our results suggest that, at least in healthy subjects, neither pvRSA nor HF power provides superior measures of RSA compared to RMSSD. This favors the RMSSD measure as the most cost-efficient measure of RSA, because it is most easily obtained with the least effort on the part of the experimenter and the lowest burden for the participant. This is particularly true in comparison to pvRSA, as this measure necessitates additional recording of the ambulatory respiration signal, which in turn requires the wearing of additional electrodes or a vest, thereby adding to the discomfort of the subjects. However, an obvious advantage of the respiration signal is that it allows a number of additional parameters to be computed from ambulatory recordings, including respiration depth and frequency (de Geus et al., 1995, 2005; Grossman, 2004; Wilhelm et al., 2003), which are important parameters in themselves. Indeed Ritz and Dahme (2006) recently argued that RSA can be used as a measure of cardiac vagal control only when taking respiratory behavior into account. Likewise, when Fourier or Wavelet analysis are used to obtain HF power, heart rate variability at a lower frequency (0.07–0.14 Hz) can be measured, which may potentially index sympathetic nervous system activity (Malliani, Pagani, Lombardi, & Cerutti, 1991). Furthermore, heart rate variability at very low frequencies (0.001–0.07) can be measured, which has been associated with an increased risk for cardiovascular disease independent of HF power (Hadase et al., 2004; La Rovere et al., 2003).

**Table 5.** *Correlations between RMSSD, pvRSA, and HF Power Separately for Lying during Sleep, Sitting, and Standing/Walking*

| | First measurement | | | | | | | | | Second measurement | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sleep | | | Sitting | | | Standing/walking | | | Sleep | | | Sitting | | | Standing/walking | | |
| | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF | RMSSD | pvRSA | HF |
| First measurement | | | | | | | | | | | | | | | | | | |
| Sleep | | | | | | | | | | | | | | | | | | |
| RMSSD | 1.00 | | | | | | | | | | | | | | | | | |
| pvRSA | .79** | 1.00 | | | | | | | | | | | | | | | | |
| HF | .94** | .83** | 1.00 | | | | | | | | | | | | | | | |
| Sitting | | | | | | | | | | | | | | | | | | |
| RMSSD | .73** | .50** | .66** | 1.00 | | | | | | | | | | | | | | |
| pvRSA | .63** | .59** | .67** | .87** | 1.00 | | | | | | | | | | | | | |
| HF | .70** | .53** | .73** | .93** | .88** | 1.00 | | | | | | | | | | | | |
| Standing/walking | | | | | | | | | | | | | | | | | | |
| RMSSD | .55** | .32** | .46** | .87** | .72** | .74** | 1.00 | | | | | | | | | | | |
| pvRSA | .52** | .46** | .51** | .76** | .84** | .74** | .76** | 1.00 | | | | | | | | | | |
| HF | .57** | .39** | .57** | .87** | .78** | .86** | .94** | .78** | 1.00 | | | | | | | | | |
| Second measurement | | | | | | | | | | | | | | | | | | |
| Sleep | | | | | | | | | | | | | | | | | | |
| RMSSD | .73** | .50** | .71** | .59** | .42** | .57** | .39** | .29* | .40** | 1.00 | | | | | | | | |
| pvRSA | .58** | .72** | .61** | .36** | .40** | .39** | .21 | .26 | .24 | .77** | 1.00 | | | | | | | |
| HF | .73** | .60** | .81** | .54** | .51** | .64** | .34* | .36** | .46** | .96** | .77** | 1.00 | | | | | | |
| Sitting | | | | | | | | | | | | | | | | | | |
| RMSSD | .53** | .24 | .51** | .70** | .65** | .70** | .48** | .46** | .52** | .70** | .53** | .65** | 1.00 | | | | | |
| pvRSA | .45** | .42** | .51** | .52** | .68** | .59** | .34** | .48** | .43** | .58** | .71** | .59** | .81** | 1.00 | | | | |
| HF | .58** | .39** | .67** | .67** | .70** | .80** | .42** | .50** | .59** | .73** | .58** | .77** | .94** | .81** | 1.00 | | | |
| Standing/walking | | | | | | | | | | | | | | | | | | |
| RMSSD | .45** | .18 | .45** | .60** | .55** | .59** | .44** | .36** | .47** | .67** | .47** | .61** | .92** | .73** | .86** | 1.00 | | |
| pvRSA | .30* | .28* | .37** | .41** | .59** | .47** | .28* | .44** | .35** | .48** | .62** | .50** | .73** | .92** | .71** | .73** | 1.00 | |
| HF | .54** | .34** | .63** | .62** | .64** | .75** | .41** | .45** | .57** | .71** | .52** | .75** | .90** | .74** | .95** | .94** | .70** | 1.00 |

*Correlation is significant at 0.05 level.
**Correlation is significant at 0.01 level.

**Table 6.** *The Fit Indices of All Tested Models*

| | | | | | Stand. residuals | |
|---|---|---|---|---|---|---|
| | | $\chi^2_{diff}$ | *df* | CFI | min | max |
| Test day (TD) | | | | | | |
| | TD | 130.12 | 72 | .98 | − .96 | 1.65 |
| Ambulatory condition (AMB) | | | | | | |
| Test day 1 | AMBa | 95.53 | 21 | .94 | − 1.78 | 1.53 |
| Test day 2 | AMBb | 73.35 | 21 | .95 | − 1.00 | 1.63 |
| Groups | | | | | | |
| | IBI | 107.54 | 72 | .97 | − 2.33 | 2.55 |
| | RR | 136.67 | 72 | .95 | − 1.92 | 1.93 |

In this study, we used the Wavelet approach to obtain HF power rather than Fourier analysis. Although Fourier analysis has been the more common approach so far, we preferred the Wavelet approach because, in contrast to Fourier analysis, it is not sensitive to violations of the stationarity assumption. Such violations are likely to occur across longer ambulatory recording periods. To deal with this, many ambulatory studies using Fourier transformation have divided the recording into smaller periods of, for example, 5 or 10 min. This reduces probability of nonstationarity and leads to a convergence of Fourier and Wavelet results (Houtveen & Molenaar, 2001). Cross-method convergence may be lower in cases of longer periods as in the current study. Therefore, it remains to be demonstrated whether the correlations of HF power to RMSSD and pvRSA also holds when Fourier-based HF estimates are used. Wavelet transformation has the additional advantage of allowing precise localization of particular RSA events in time by providing a full time-frequency decomposition of the IBI time series (e.g., see Pichot et al., 1999). In the current study, we have not used this additional time-frequency information because no comparable information can be obtained from RMSSD and pvRSA.

A limitation of this study that is worth mentioning is that we examined the relative behavior of the three ambulatory RSA measures by comparing them to each other across time and ambulatory conditions. True validity testing, however, would require comparison of each of the RSA measures to some future cardiovascular disease endpoint or to an external validation criterion of cardiac vagal control, which would be the most plausible explanation for any cardioprotective effects associated with high levels of RSA. Although a number of studies have shown predictive validity of RMSSD (Dekker et al., 2000; Nolan et al., 1998; Singh et al., 1998; Tsuji et al., 1996) and HF power (Bigger et al., 1993; Singh et al., 1998; Tsuji et al., 1996) for cardiovascular disease, none has done so for pvRSA. However, in these prospective studies, HF and RMSSD were either measured during short periods (Dekker et al., 2000; Singh et al., 1998; Tsuji et al., 1996) or when full 24-h ambulatory recording was used in patient samples (Bigger et al., 1993; Nolan et al., 1998). Whether prolonged recording of RSA in naturalistic settings has predictive power in the population at large remains to be established. Taking into account the high correlations among the three measures in this study, it is hard to imagine that one of these measures would exceed the others in predictive power.

To test which of these RSA measures is most closely associated with individual differences in cardiac vagal control, ambulatory recordings could be made under partial and full parasympathetic blockade. We do not consider such long-term (i.e., 24-h) pharmacological interventions feasible in a true naturalistic setting. However, a number of studies in controlled experimental settings have shown that RMSSD, pvRSA, and HF power all respond to muscarinergic blockade by showing a graded, almost linear, decrease with increasing dose (Berntson et al., 1997; Cacioppo et al., 1994; Martinmaki et al., 2006; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Combined, these findings and the high correlations among the three measures in a realistic ambulatory setting render it unlikely that one of these measures would exceed the others in detecting individual differences in daily cardiac vagal control. In summary, we conclude that ambulatory RMSSD, pvRSA, and HF power are highly correlated and that their correlation is stable across time, ambulatory conditions, and a wide range of resting RR and HR values. Because the different RSA measurement strategies have varying specific advantages, for instance, providing additional information on RR or on (very) low frequency power, the choice for a specific measure should be based on the exact research questions. In large-scale research that focuses entirely on individual differences in RSA as correlates or predictors of disease risk, RMSSD appears to be the most cost-efficient measure.

## REFERENCES


Akselrod, S., Gordon, D., Madwed, J. B., Snidman, N. C., Shannon, D. C., & Cohen, R. J. (1985). Hemodynamic regulation: investigation by spectral analysis. *American Journal of Physiology*, *249*, H867–H875.

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. C., & Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science*, *213*, 220–222.

Allen, M. T., & Crowell, M. D. (1989). Patterns of autonomic response during laboratory stressors. *Psychophysiology*, *26*, 603–614.

Ando, M., Katare, R. G., Kakinuma, Y., Zhang, D., Yamasaki, F., Muramoto, K., et al. (2005). Efferent vagal nerve stimulation protects heart against ischemia-induced arrhythmias by preserving connexin43 protein. *Circulation*, *112*, 164–170.

Berntson, G. G., Bigger, J. T. Jr., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*, 623–648.

Berntson, G. G., Lozano, D. L., & Chen, Y. J. (2005). Filter properties of root mean square successive difference (RMSSD) for heart rate. *Psychophysiology*, *42*, 246–252.

Bigger, J. T. Jr., Fleiss, J. L., Rolnitzky, L. M., & Steinman, R. C. (1992). Stability over time of heart period variability in patients with previous myocardial infarction and ventricular arrhythmias. The CAPS and ESVEM investigators. *American Journal of Cardiology*, *69*, 718–723.

Bigger, J. T. Jr., Fleiss, J. L., Rolnitzky, L. M., & Steinman, R. C. (1993). Frequency domain measures of heart period variability to assess risk late after myocardial infarction. *Journal of the American College of Cardiology*, *21*, 729–736.

Bigger, J. T. Jr., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Correlations among time and frequency domain measures of heart period variability two weeks after acute myocardial infarction. *American Journal of Cardiology*, *69*, 891–898.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Boomsma, D. I., Beem, A. L., van den Berg, M., Dolan, C. V., Koopmans, J. R., Vink, J. M., et al. (2000). Netherlands twin family study of anxious depression (NETSAD). *Twin Research*, *3*, 323–334.

Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*, 403–421.

Byrne, E. A., & Porges, S. W. (1993). Data-dependent filter characteristics of peak-valley respiratory sinus arrhythmia estimation: A cautionary note. *Psychophysiology*, *30*, 397–404.

Cacioppo, J. T., Berntson, G. G., Binkley, P. F., Quigley, K. S., Uchino, B. N., & Fieldstone, A. (1994). Autonomic cardiac control. II. Noninvasive indices and basal response as revealed by autonomic blockades. *Psychophysiology*, *31*, 586–598.

de Geus, E. J., Posthuma, D., Kupper, N., van den Berg, M., Willemsen, G., Beem, A. L., et al. (2005). A whole-genome scan for 24-hour respiration rate: A major locus at 10q26 influences respiration during sleep. *American Journal of Human Genetics*, *76*, 100–111.

de Geus, E. J., Willemsen, G. H., Klaver, C. H., & van Doornen, L. J. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biological Psychology*, *41*, 205–227.

Dekker, J. M., Crow, R. S., Folsom, A. R., Hannan, P. J., Liao, D., Swenne, C. A., et al. (2000). Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: The ARIC study. *Circulation*, *102*, 1239–1244.

Dekker, J. M., Schouten, E. G., Klootwijk, P., Pool, J., Swenne, C. A., & Kromhout, D. (1997). Heart rate variability from short electrocardiographic recordings predicts mortality from all causes in middle-aged and elderly men. The Zutphen Study. *American Journal of Epidemiology*, *145*, 899–908.

Gerin, W., Rosofsky, M., Pieper, C., & Pickering, T. G. (1994). A test of generalizability of cardiovascular reactivity using a controlled ambulatory procedure. *Psychosomatic Medicine*, *56*, 360–368.

Goldstein, I. B., Shapiro, D., & Guthrie, D. (2006). Ambulatory blood pressure and family history of hypertension in healthy men and women. *American Journal of Hypertension*, *19*, 486–491.

Grossman, P. (1992). Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: Selection of window frames, keeping the panes clean and viewing the neural topography. *Biological Psychology*, *34*, 131–161.

Grossman, P. (2004). The LifeShirt: A multi-function ambulatory system monitoring health, disease, and medical intervention in the real world. *Studies in Health Technology and Informatics*, *108*, 133–141.

Grossman, P., Karemaker, J., & Wieling, W. (1991). Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: The need for respiratory control. *Psychophysiology*, *28*, 201–216.

Grossman, P., & Kollai, M. (1993). Respiratory sinus arrhythmia, cardiac vagal tone, and respiration: Within- and between-individual relations. *Psychophysiology*, *30*, 486–495.

Grossman, P., van Beek, J., & Wientjes, C. (1990). A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. *Psychophysiology*, *27*, 702–714.

Grossman, P., Wilhelm, F. H., & Spoerle, M. (2004). Respiratory sinus arrhythmia, cardiac vagal control, and daily activity. *American Journal of Physiology. Heart and Circulatory Physiology*, *287*, H728–H734.

Hadase, M., Azuma, A., Zen, K., Asada, S., Kawasaki, T., Kamitani, T., et al. (2004). Very low frequency power of heart rate variability is a powerful predictor of clinical prognosis in patients with congestive heart failure. *Circulation Journal*, *68*, 343–347.

Hatfield, B. D., Spalding, T. W., Santa Maria, D. L., Porges, S. W., Potts, J. T., Byrne, E. A., et al. (1998). Respiratory sinus arrhythmia during exercise in aerobically trained and untrained men. *Medicine and Science in Sports and Exercise*, *30*, 206–214.

Hayano, J., Sakakibara, Y., Yamada, A., Yamada, M., Mukai, S., Fujinami, T., et al. (1991). Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects. *American Journal of Cardiology*, *67*, 199–204.

Hohnloser, S. H., Klingenheben, T., Zabel, M., Schroder, F., & Just, H. (1992). Intraindividual reproducibility of heart rate variability. *Pacing and Clinical Electrophysiology*, *15*, 2211–2214.

Houtveen, J. H., Groot, P. F., & de Geus, E. J. (2005). Effects of variation in posture and respiration on RSA and pre-ejection period. *Psychophysiology*, *42*, 713–719.

Houtveen, J. H., & Molenaar, P. C. (2001). Comparison between the Fourier and Wavelet methods of spectral analysis applied to stationary and nonstationary heart period data. *Psychophysiology*, *38*, 729–735.

Houtveen, J. H., Rietveld, S., & de Geus, E. J. (2002). Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise. *Psychophysiology*, *39*, 427–436.

Hull, S. S. Jr., Evans, A. R., Vanoli, E., Adamson, P. B., Stramba-Badiale, M., Albert, D. E., et al. (1990). Heart rate variability before and after myocardial infarction in conscious dogs at high and low risk of sudden death. *Journal of the American College of Cardiology*, *16*, 978–985.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.50: User's reference guide*. Chicago: Scientific Software International.

Kamarck, T. W., Schwartz, J. E., Janicki, D. L., Shiffman, S., & Raynor, D. A. (2003). Correspondence between laboratory and ambulatory measures of cardiovascular reactivity: A multilevel modeling approach. *Psychophysiology*, *40*, 675–683.

Kamphuis, A., & Frowein, H. W. (1985). Assessment of mental effort by means of heart rate spectral analysis. In J. F. Orlebeke, L. J. P. van Doornen, & G. Mulder (Eds.), *The psychophysiology of cardiovascular control*. New York: Plenum Press.

Katona, P. G., & Jih, F. (1975). Respiratory sinus arrhythmia: Noninvasive measure of parasympathetic cardiac control. *Journal of Applied Physiology*, *39*, 801–805.

Kleiger, R. E., Bigger, J. T., Bosner, M. S., Chung, M. K., Cook, J. R., Rolnitzky, L. M., et al. (1991). Stability over time of variables measuring heart rate variability in normal subjects. *American Journal of Cardiology*, *68*, 626–630.

Kupper, N. H., Willemsen, G., van den Berg, M., de Boer, D., Posthuma, D., Boomsma, D. I., et al. (2004). Heritability of ambulatory heart rate variability. *Circulation*, *110*, 2792–2796.

La Rovere, M. T., Pinna, G. D., Maestri, R., Mortara, A., Capomolla, S., Febo, O., et al. (2003). Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients. *Circulation*, *107*, 565–570.

Langewitz, W., & Ruddel, H. (1989). Spectral analysis of heart rate variability under mental stress. *Journal of Hypertension*, *7*(Suppl.), S32–S33.

Levy, M. N., & Schwartz, P. J. (1994). *Vagal control of the heart: Experimental basis and clinical implications*. Armonk, NY: Futura Publishing Co.

Lombardi, F., Sandrone, G., Pernpruner, S., Sala, R., Garimoldi, M., Cerutti, S., et al. (1987). Heart rate variability as an index of sympathovagal interaction after acute myocardial infarction. *American Journal of Cardiology*, *60*, 1239–1245.

Malliani, A., Pagani, M., Lombardi, F., & Cerutti, S. (1991). Cardiovascular neural regulation explored in the frequency domain. *Circulation*, *84*, 482–492.

Martinmaki, K., Rusko, H., Kooistra, L., Kettunen, J., & Saalasti, S. (2006). Intraindividual validation of heart rate variability indexes to measure vagal effects on hearts. *American Journal of Physiology. Heart and Circulatory Physiology*, *290*, H640–H647.

Mulder, L. J. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, *34*, 205–236.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus User's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Nolan, J., Batin, P. D., Andrews, R., Lindsay, S. J., Brooksby, P., Mullen, M., et al. (1998). Prospective study of heart rate variability and mortality in chronic heart failure: Results of the United Kingdom Heart Failure Evaluation and Assessment of Risk Trial (UK-Heart). *Circulation*, *98*, 1510–1516.

Penttila, J., Helminen, A., Jartti, T., Kuusela, T., Huikuri, H. V., Tulppo, M. P., et al. (2001). Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: Effects of various respiratory patterns. *Clinical Physiology*, *21*, 365–376.

Pichot, V., Gaspoz, J. M., Molliex, S., Antoniadis, A., Busso, T., Roche, F., et al. (1999). Wavelet transform to quantify heart rate variability and

to assess its instantaneous changes. *Journal of Applied Physiology*, *86*, 1081–1091.

Pickering, T. G., & Devereux, R. B. (1987). Ambulatory monitoring of blood pressure as a predictor of cardiovascular risk. *American Heart Journal*, *114*, 925–928.

Pitzalis, M. V., Mastropasqua, F., Massari, F., Forleo, C., Di Maggio, M., Passantino, A., et al. (1996). Short- and long-term reproducibility of time and frequency domain heart rate variability measurements in normal subjects. *Cardiovascular Research*, *32*, 226–233.

Riese, H., Groot, P. F., van den Berg, M., Kupper, N. H., Magnee, E. H., Rohaan, E. J., et al. (2003). Large-scale ensemble averaging of ambulatory impedance cardiograms. *Behavior Research Methods, Instruments, & Computers*, *35*, 467–477.

Ritz, T., & Dahme, B. (2006). Implementation and interpretation of respiratory sinus arrhythmia measures in psychosomatic medicine: Practice against better evidence? *Psychosomatic Medicine*, *68*, 617–627.

Sakakibara, M., Takeuchi, S., & Hayano, J. (1994). Effect of relaxation training on cardiac parasympathetic tone. *Psychophysiology*, *31*, 223–228.

Saul, J. P., Arai, Y., Berger, R. D., Lilly, L. S., Colucci, W. S., & Cohen, R. J. (1988). Assessment of autonomic regulation in chronic congestive heart failure by heart rate spectral analysis. *American Journal of Cardiology*, *61*, 1292–1299.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23–74.

Singer, D. H., Martin, G. J., Magid, N., Weiss, J. S., Schaad, J. W., Kehoe, R., et al. (1988). Low heart rate variability and sudden cardiac death. *Journal of Electrocardiology*, *21*(Suppl.), S46–S55.

Singh, J. P., Larson, M. G., Tsuji, H., Evans, J. C., O'Donnell, C. J., & Levy, D. (1998). Reduced heart rate variability and new-onset hypertension: Insights into pathogenesis of hypertension: The Framingham Heart Study. *Hypertension*, *32*, 293–297.

Sinnreich, R., Kark, J. D., Friedlander, Y., Sapoznikov, D., & Luria, M. H. (1998). Five minute recordings of heart rate variability for population studies: Repeatability and age–sex characteristics. *Heart*, *80*, 156–162.

Stein, P. K., Rich, M. W., Rottman, J. N., & Kleiger, R. E. (1995). Stability of index of heart rate variability in patients with congestive heart failure. *American Heart Journal*, *129*, 975–981.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, *93*, 1043–1065.

Tsuji, H., Larson, M. G., Venditti, F. J. Jr., Manders, E. S., Evans, J. C., Feldman, C. L., et al. (1996). Impact of reduced heart rate variability on risk for cardiac events. The Framingham Heart Study. *Circulation*, *94*, 2850–2855.

Tulppo, M. P., Makikallio, T. H., Takala, T. E., Seppanen, T., & Huikuri, H. V. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American Journal of Physiology*, *271*, H244–H252.

Van Doornen, L. J. P., Knol, D., Willemsen, A. H. M., & de Geus, E. J. C. (1994). The relationship between stress reactivity in the lab and in real-life: Is reliability the limiting factor? *Journal of Psychophysiology*, *8*, 297–304.

Vanoli, E., De Ferrari, G. M., Stramba-Badiale, M., Hull, S. S. Jr., Foreman, R. D., & Schwartz, P. J. (1991). Vagal stimulation and prevention of sudden death in conscious dogs with a healed myocardial infarction. *Circulation Research*, *68*, 1471–1481.

Verdecchia, P., Porcellati, C., Schillaci, G., Borgioni, C., Ciucci, A., Battistelli, M., et al. (1994). Ambulatory blood pressure. An independent predictor of prognosis in essential hypertension. *Hypertension*, *24*, 793–801.

Verdecchia, P., Schillaci, G., Borgioni, C., Ciucci, A., Pede, S., & Porcellati, C. (1998). Ambulatory pulse pressure: A potent predictor of total cardiovascular risk in hypertension. *Hypertension*, *32*, 983–988.

Verdecchia, P., Schillaci, G., Reboldi, G., Franklin, S. S., & Porcellati, C. (2001). Ambulatory monitoring for prediction of cardiac and cerebral events. *Blood Pressure Monitoring*, *6*, 211–215.

Vrijkotte, T. G., van Doornen, L. J., & de Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*, 880–886.

Vrijkotte, T. G. M., Riese, H., & de Geus, E. J. C. (2001). Cardiovascular reactivity to work stress assessed by ambulatory blood pressue, heart rate, and heart rate variability. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment. Computer assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 345–360). Seattle: Hogrefe & Huber.

Weber, E. J., Molenaar, P. C., & van der Molen, M. W. (1992). A nonstationarity test for the spectral analysis of physiological time series with an application to respiratory sinus arrhythmia. *Psychophysiology*, *29*, 55–65.

Wiklund, U., Akay, M., & Niklasson, U. (1997). Short-term analysis of heart-rate variability by adapted wavelet transforms. *IEEE Engineering in Medicine and Biology Magazine*, *16*, 113–8, 138.

Wilhelm, F. H., Roth, W. T., & Sackner, M. A. (2003). The LifeShirt. An advanced system for ambulatory measurement of respiratory and cardiac function. *Behavior Modification*, *27*, 671–691.

Willemsen, G. H., de Geus, E. J., Klaver, C. H., van Doornen, L. J., & Carroll, D. (1996). Ambulatory monitoring of the impedance cardiogram. *Psychophysiology*, *33*, 184–193.

(Received July 19, 2006; Accepted November 15, 2006)