SHORT COMMUNICATION

Two-week test–retest reliability of the Polar® RS800CX™ to record heart rate variability

DeWayne P. Williams¹, Marc N. Jarczok², Robert J. Ellis³, Thomas K. Hillecke⁴, Julian F. Thayer¹ and Julian Koenig¹,⁵

¹Department of Psychology, The Ohio State University, Columbus, OH, USA, ²Mannheim Institute of Public Health, Social and Preventive Medicine, Mannheim Medical Faculty, Heidelberg University, Mannheim, Germany, ³School of Computing, National University of Singapore, Singapore City, Malaysia, ⁴School of Therapeutic Sciences, SRH University of Applied Sciences, and ⁵Section for Translational Psychobiology in Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre for Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany

Summary

Recently, research has validated the use of Polar® heart rate monitors as a tool to index heart rate variability (HRV). In the current investigation, we sought to evaluate the test–retest reliability of both time and frequency domain measures of HRV using the Polar® RS800CX™. Continuous HRV data were collected as 60 nominally healthy adults underwent a resting and orthostatic stress test. We evaluated reproducibility by means of the interclass correlation coefficient for absolute agreement and consistency, and the standard error of measurement. We found moderate reliable 2-week test–retest reliability of HRV using the Polar® RS800CX™, results that are in line with previous studies that have validated the stability of HRV using other methods of measurement (e.g. electrocardiogram). Additionally, when examining different methods of spectral density estimation, we found that using the auto-regressive transformation method provides the most stable indices of HRV. Taken together, our results suggest that the Polar® RS800CX™ is not only a valid method to record HRV, but also a reliable one, particularly when using the auto-regressive transformation method.

Correspondence

Julian Koenig, Section for Translational Psychobiology in Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre for Psychosocial Medicine, University of Heidelberg, Blumenstrasse 8, 69115 Heidelberg, Germany
E-mail: Julian.Koenig@med.uni-heidelberg.de

Research conducted at the School of Therapeutic Sciences, SRH University of Applied Sciences, Heidelberg, Germany.

Accepted for publication
Received 27 July 2015; accepted 12 October 2015

Key words
stability; orthostatic stress; spectral analysis; parasympathetic nervous system; sympathetic nervous system; autonomic nervous system; heart rate variability

Introduction

Heart rate variability (HRV) refers to the beat-to-beat fluctuations in heart period governed by both the sympathetic (SNS) and parasympathetic (PNS) branches of the autonomic nervous system (ANS). The basic data for the calculation of all the measures of HRV is the sequence of time intervals between adjacent heartbeats – the interbeat interval (IBI). The consensus is that an electrocardiogram (ECG) is preferred to index IBIs and derive HRV; however, an ECG can be expensive and difficult to use in ambulatory research and other settings.

Therefore, a number of portable devices have been created to measure both heart rate (HR) and IBIs without the use of an ECG. Among these series of devices are Polar® HR monitors, that have been used in a number of empirical investigations that measured HRV under various conditions. A number of previous studies have shown the validity, reliability and effectiveness of such devices in comparison with ECG methods (Kingsley et al., 2004; Gamelin et al., 2006; Barbosa et al., 2014). However, to our knowledge, research has yet to examine the test–retest reliability Polar® HR monitors, in addition to the evaluation of both time domain and frequency domain measures of HRV when using Polar® HR monitors. Time domain measures range from short-term (e.g. the standard deviation of IBIs or the root-mean-square successive differences in an IBI series within a 5-min window) to long-term (e.g. the standard deviation of all IBIs in a 24-h window) periods. Frequency domain measures (e.g. Berntson et al., 1997; Task Force, 1996; for detailed methodological discussions, see Porges & Bohrer, 1990 and Tarvainen & Niskanen, 2008) quantify HRV from an IBI time series that has been detrended (to remove slow nonstationarities) using a moving
polynomial filter (such as a cubic spline; Porges & Bohrer, 1990) or a smoothness priors regularization (Tarvainen & Niskanen, 2008). The detrended IBI time series is then decomposed into its underlying periodicities, and a power spectrum density plot is created, plotting spectral power density (in ms$^2$ or s$^2$) as a function of frequency (in Hz).

Two common solutions are used: a nonparametric fast Fourier transform (FFT) and a parametric autoregressive algorithm (AR) (e.g. Kay & Marple, 1981). The FFT algorithm utilizes Welch’s periodogram method. This divides the sample into 256-ms windows that overlap by 50%, and averages overlapping segments. This decreases the variance of the FFT spectrum. Absolute power values are then obtained by integrating the spectrum within two prespecified frequency bands (Fig. 1a). Because of the fast breakdown of acetylcholine, PNS modulation of the heart is fast and short-lived. Thus, power in the high-frequency (HF) band (0.15–0.4 Hz) is regarded as the direct and exclusive consequence of PNS activity. Activity in the low-frequency (LF) band (0.04–0.15 Hz) is considered to reflect joint activation of the PNS and SNS (cf. Task Force, 1996).

The AR algorithm uses a factorization procedure to obtain a distinct LF and HF frequency component (Fig. 1b). Power values are obtained as the powers of those components. The advantages of an AR solution are smoother spectral components that are independent of prespecified frequency bands, clear central frequencies of each component, and an accurate estimation of power spectral density even on a small number of (stationary) samples (Task Force, 1996). Furthermore, the central frequency of the HF component has been shown to serve as an index of respiration rate (i.e. frequency in Hz × 60 = RR; Thayer et al., 2002).

A recent study (Dantas et al., 2010) compared the reproducibility of time domain and spectral HRV indexes of a group of healthy individuals in the supine and standing positions using short-term recording periods under spontaneous breathing. Most HRV measures were reproducible independent of body position, but better correlation indexes were obtained in the orthostatic position. The authors concluded that increased SNS activity during the orthostatic (standing) position facilitates reproducibility of the main HRV indexes. However, the authors investigated the reproducibility of two assessments only two hours apart. To our knowledge, no study investigated reliability of HRV measures in healthy subjects during simple cardiovascular reflex tests under free breathing conditions across a longer time interval, particularly using mobile devices such as Polar® HR monitors that are often used in both psychological and physiological studies. In this study, we aim to investigate the general consistency of different time and frequency domain measures of HRV within a 2-week interval in subjects undergoing an orthostatic stress test using Polar® RS800CX™ (Polar Electro Inc., Lake Success, NY, USA) HR monitors. In particular, we highlight the impact of different methods of parametric and nonparametric spectral analyses (FFT and AR, respectively) on the reliability of the frequency domain measures of HRV using Polar® mobile devices.

Material and methods

Subjects

We studied 60 nominally healthy adults (44 female; mean age: 23.7; SD: 4.2, range: 19–45 years) recruited at the SRH University Heidelberg (Germany) by study advertisement. Self-rated health (SRH) was measured using the question ‘How do you rate your current health status?’ on a 0 ‘very bad’ to 6 ‘excellent’ scale. Only subjects indicating a SRH ≥ 3 ‘fair’ were included in the trial. None of the subjects reported current medication intake or suffering from chronic or acute disease such as diabetes or cardiovascular disease. The study was approved by the local Ethics Committee, and all participants provided written informed consent prior to their assessment. The investigation conformed to the principles outlined in the Declaration of Helsinki.

Experiment

Each participant was scheduled on 2 weekdays; exactly 2 weeks apart at the same time of day (9 a.m. to 6 p.m.). Date and time of measurements were recorded by a protocol. All participants received class credits or an allowance of 20€ for completion of the study. Assessments took place during September 2012 and January 2013. HRV was measured within a 15-min orthostatic stress test (5 min each sitting in a chair, standing and sitting again). A Polar® RS800CX™ HR monitor was used to assess IBIs. This device uses a transmitter consisting of a stable polyamide case with electrodes attached to an elastic belt fixated to the chest.
of the subjects was used to record IBIs at a sampling frequency of 1000 Hz, providing a temporal resolution of 1 ms for each R–R interval. This data is simultaneously transmitted to the watch, and Polar® ProTrainer 5 was used to transfer recordings from the watch onto a personal computer. IBI data were imported into ‘Kubios HRV’ (Biosignal Analysis and Medical Imaging Group, University Kuopio, Finland, version 2.0, Tarvainen et al., 2009) for analysis. The following time domain measures were derived: mean IBI (ms), the standard deviation of normal to normal intervals (SDNN, ms), the square root of the mean squared difference of successive NN intervals (RMSSD, ms), the number of pairs of successive NNs that differ by more than 50 ms (NN50, count) and the proportion of NN50 divided by total number of NN (pNN50, %) components. Both non-parametric (FFT) and parametric (AR) frequency domain models were performed, and the following measures calculated: high frequency power (HF; 0.15–0.4 Hz), low frequency power (LF; 0.04–0.15 Hz) power, their relative power (LF%, HF%) and absolute spectral power expressed as normalized units (LF n.u., HF n.u.), and the LF/HF ratio (Task Force, 1996).

**Calculation**

Kubios output was imported in SPSS (ver. 21; IBM, Chicago, IL, USA) for further analysis. Test–retest reliability for HRV measures was assessed via a three-layered approach as recommended by Weir (2005). First, repeated-measures two-way ANOVA (HRV measure × time point) was performed as described to examine systematic error. Normal distribution of variables was assessed using the Shapiro–Wilk test. If data were skewed (and thus not meeting the assumptions for ANOVA), a Wilcoxon–Mann–Whitney test was used instead. Second, Pearson product–moment correlations were expressed by the interclass correlation coefficient (ICC, with values ranging from 0 to 1) and 95% confidence interval for absolute agreement between the two assessments. The following interpretation of ICC was used: ≤0.2 indicates poor agreement; 0.3–0.4 indicates fair agreement; 0.5–0.6 indicates moderate agreement; 0.7–0.8 indicates strong agreement; and >0.8 indicates almost perfect agreement (Weir, 2005). Third, the standard error of measurement (SEM = SD × √(1−ICC)) was calculated. For all tests, a P-value <0.05 was taken as statistically significant. Furthermore, to address the reliability of measurement for the two different approaches for HRV spectral power estimation (FFT or AR), ICCs and SEMs were calculated for each of the methods applied.

**Results**

Descriptive statistics of all raw HRV parameters measured during the orthostatic stress test in the two assessments taken 2 weeks apart from each other are given in Table 1.

Reliability indices for HRV measures for the single conditions of the orthostatic stress test were calculated. Statistical analysis suggests the absence of a systematic change, as revealed by analysis of variance (Table 1). ICCs and SEMs for time domain measures and frequency domain measures were calculated on raw data (Table 2). ICCs for absolute agreement were calculated for each condition and revealed poor-to-moderate reliability of HRV measurements. Concerning the 2-week reliability of time domain measures, we found moderate (ICCs: 0.5–0.6) agreement in almost every time measure independent of the condition of the orthostatic stress test. NN50 (ICCs: 0.59–0.69) and pNN50 (ICC: 0.67–0.73) components seem to be most reliable among the time domain measures, as indexed by the highest ICC. However, SDNN showed the poorest reliability as index by the lowest ICC. Overall, ICCs showed greater variance among the frequency domain measures, revealing that the frequency band (LF or HF), the method of frequency estimation (FFT or AR) and data treatment (e.g. normalized units) have a noticeable impact on the reliability of the selected measure.

**Discussion**

**Two-week reliability of HRV measurements using Polar® HR monitors**

The reliability of HRV measures within repeated-measures designs is of interest for clinical and experimental research alike. Two previous studies have addressed the reliability of short-term HRV measures during simple cardiovascular reflex tests using relatively small data sets (Jauregui-Renaud et al., 2001; Dantas et al., 2010). The present paper reports the largest data set addressing the test–retest reliability of HRV measures during different orthostatic conditions. Consistently with previous literature, our results support the notion that HRV is a moderately reliable measurement (Sandercock et al., 2005) in unadjusted models. However, testing the reliability using the Polar® RS800CX™ device suggests that the mobile HR devices show good stability over a 2-week interval. Moreover, our results provide the most stable method of deriving HRV, particularly when using Polar® devices.

**Differences among orthostatic stress condition**

The method of relative frequency power calculation (%) shows different reliability between the orthostatic stress condition and the method of frequency domain estimation (FFT or AR). Interestingly, while relative HF power shows poor reliability independent of the condition of measurement, relative LF power shows poor reliability during sitting (1) and standing, and only moderate reliability during the sitting (2) condition, independently of the method of frequency domain estimation. In general, our data supports previous findings that the LF component is more reliable under free breathing conditions than the HF component (Pitzalis et al., 1996).

**Methods of frequency estimation**

The reproducibility of HRV measures depends on different analytical procedures, especially for the computation of fre-
Table 1  Descriptive results for HRV parameters during orthostatic stress test in the two tests taken 2 weeks apart from each other.

<table>
<thead>
<tr>
<th>Orthostatic stress test</th>
<th>Measurement 1 (T₁)</th>
<th>Measurement 2 (T₂)</th>
<th>Difference (T₂-T₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sitting (s)</td>
<td>Standing</td>
<td>Sitting (s)</td>
</tr>
<tr>
<td>Mean RR (ms)</td>
<td>775.88</td>
<td>677.382 (95.06)</td>
<td>765.87</td>
</tr>
<tr>
<td>SDNN (ms)</td>
<td>93.94</td>
<td>64.77 (22.03)</td>
<td>93.74 (48.71)</td>
</tr>
<tr>
<td>RMSSD (ms)</td>
<td>72.20</td>
<td>33.50 (16.67)</td>
<td>61.12 (43.39)</td>
</tr>
<tr>
<td>NN50 (count)</td>
<td>82.07</td>
<td>34.50 (31.58)</td>
<td>75.37 (51.99)</td>
</tr>
<tr>
<td>pNN50 (%)</td>
<td>22.59</td>
<td>8.34 (8.40)</td>
<td>20.47 (16.11)</td>
</tr>
<tr>
<td>HF FFT (%)</td>
<td>21.26</td>
<td>11.19 (7.97)</td>
<td>21.92 (17.57)</td>
</tr>
<tr>
<td>LF FFT (%)</td>
<td>31.69</td>
<td>38.26 (14.89)</td>
<td>29.94 (15.94)</td>
</tr>
<tr>
<td>HF FFT (n.u.)</td>
<td>39.44</td>
<td>22.06 (13.09)</td>
<td>40.68 (19.78)</td>
</tr>
<tr>
<td>LF FFT (n.u.)</td>
<td>60.56</td>
<td>77.94 (13.09)</td>
<td>59.32 (19.78)</td>
</tr>
<tr>
<td>HF AR (%)</td>
<td>18.36</td>
<td>10.29 (6.80)</td>
<td>18.74 (15.99)</td>
</tr>
<tr>
<td>LF AR (%)</td>
<td>32.57</td>
<td>36.48 (14.58)</td>
<td>28.59 (14.01)</td>
</tr>
<tr>
<td>HF AR (n.u.)</td>
<td>35.98</td>
<td>21.79 (10.46)</td>
<td>37.84 (17.94)</td>
</tr>
<tr>
<td>LF AR (n.u.)</td>
<td>64.02</td>
<td>78.21 (10.46)</td>
<td>62.16 (17.94)</td>
</tr>
<tr>
<td>LF/HF ratio FFT</td>
<td>2.20</td>
<td>4.98 (3.18)</td>
<td>2.18 (1.91)</td>
</tr>
<tr>
<td>LF/HF ratio AR</td>
<td>2.32</td>
<td>4.81 (3.26)</td>
<td>2.30 (1.77)</td>
</tr>
</tbody>
</table>

All values are mean and standard deviation in brackets (SD).
*Significance difference on the <0.05 level (repeated-measures ANOVA or Wilcoxon–Mann–Whitney test represented by P values).
frequency domain measures. As in previous studies (Pitzalis et al., 1996), we investigated the reproducibility of the frequency domain parameters obtained from short-term recordings using both FFT and AR spectral modelling approaches. ICCs indicate that AR modelling is slightly more reliable than FFT modelling analysis comparing ICCs on relative frequency power (%). Normalized units (n.u.) show greater ICCs among the conditions independent of the method of frequency estimation. However, while the majority of ICCs for HF FFT (n.u.), LF FFT (n.u.), HF AR (n.u.) and HF AR (n.u.) show fair reliability, AR during the standing condition and FFT during the sitting (2) condition show moderate reliability. The spectral modelling method also shows considerable differences regarding the calculation of the LF/HF ratio. While FFT seems to be more reliable (in terms of larger ICCs) to AR considering the reliability of LF/HF ratio calculation during sitting (1) and sitting (2) condition, AR is more reliable during the standing condition. Pitzalis et al. found that the reproducibility of the parameters obtained using either the FFT or the AR method was generally similar. Considering baseline recordings the authors report difference within the ICCs between FFT and AR method of 0.07 for LF and 0.33 for HF, with higher reliability of the FFT method (Pitzalis et al., 1996). However, our results draw attention to an interaction of the selected HRV measures, conditions of measurement and the appropriate method of frequency estimation that researchers should be aware of. Furthermore, our results are consistent with previous findings (Badolini et al., 1998; Chemla et al., 2005; Pichon et al., 2006) that highlighted a large discrepancy between the results of FFT and AR analysis and that these parametric and nonparametric spectral analyses should not be considered as interchangeable in healthy (Pichon et al., 2006) or clinical (Chemla et al., 2005) samples, even if they give same qualitative results.

**Conclusion**

Consistent with previous findings, our results suggest that HRV is a moderately reliable measurement within a 2-week interval in healthy human adults. However, these results extend previous research, showing that this test–retest reliability is good without the use of traditional ECG methods and instead, using the Polar® RS800CX™ mobile HR monitor. These results can give researchers who use Polar® devices more confidence that Polar® derived HRV is stable and likely reflects trait influence. However, test–retest reliability varies as a function of body position during an orthostatic stress test. Furthermore, these results draw attention to the appropriate selection of methods for frequency estimation as they result in differences regarding their reliability. While stability is good using most methods to derive HRV as indicated in this report, our results support the use of AR modelling when using the Polar® RS800CX™ to collect HRV data across changing body positions.

**Acknowledgments**

The authors would like to thank all participants of the study and Claudia Bach for her assistance in data collection. JK is supported by a Physician-Scientist-Fellowship provided by the Medical School, University of Heidelberg, Germany.

**Conflict of interest**

The authors report no conflicts.
References


