Repeatability of Heart Rate Variability Measures

Emily B. Schroeder, PhD,* Eric A. Whitsel, MD, MPH,†‡ Gregory W. Evans, MA,§ Ronald J. Prineas, MD, PhD,§ Lloyd E. Chambless, PhD,¶ and Gerardo Heiss, MD, PhD†

Abstract: Due to the sparse data on the repeatability of short and ultra-short term heart rate variability (HRV) measures, we measured the repeatability of common HRV measures derived from 10-second, 2-minute, and 6-minute recordings in 63 healthy men and women, aged 45-64, in Chapel Hill, North Carolina. Three 10-second and 2 six-minute heart rate recordings were obtained during each of 2 visits, separated by 1 to 2 weeks. We partitioned the measurement error into components and computed intraclass correlation coefficients using nested, random effects models. Repeatability improved with the length of recording: intraclass correlation coefficients were greater than 0.7 for 6-minute measures and 2-minute time domain measures and greater than 0.5 for 2-minute frequency domain measures. Repeatability of measures from 10-second records was lower, but improved considerably when the mean from 2 or 3 records was used. Correlations between the same measures from different length recordings were quite high. Our findings support the use of records of at least 5 minutes in length in epidemiological studies, in accordance with previous guidelines. Researchers using 10-second records should consider taking the mean of several recordings, when possible, or using statistical methods to correct for measurement error. **Key words:** Autonomic nervous system, heart rate variability, reliability, repeatability.

The measurement of heart rate variability (HRV), the oscillation in the interval between consecutive heart beats, is an established, noninvasive and quantitative method of assessing cardiac autonomic activity. Experiments in animals and humans have characterized many of the physiological phenomena underlying various HRV measures. Epidemiologic studies have described some of the population correlates of these measures and have demonstrated that low HRV predicts post-myocardial infarction mortality (1–5), incident coronary heart disease (6–8), incident hypertension (9,10), and all cause mortality (11–13).

Historically, HRV has been measured using both 24-hour Holter recordings and shorter recordings, most commonly between 2- and 15-minutes in

*From the *Duke University School of Medicine, Durham, NC; †Departments of Epidemiology, ‡Medicine, and §Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; and the ¶Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC.*

This work was supported in part by a grant (RR00046) from the General Clinical Research Centers program of the Division of Research Resources, National Institutes of Health, and by grants (N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021 and N01-HC-55022) from the US National Heart, Lung, and Blood Institute. E.B.S. and E.A.W. were supported by grant (5-T32-HL-07055) from the National Institutes of Health.

Reprint requests: Gerardo Heiss, MD, PhD, Department of Epidemiology, University of North Carolina at Chapel Hill, 137 E Franklin St, Suite 306, Chapel Hill, NC, 27514; e-mail: gerardo_heiss@unc.edu.

0022-0736/04/3703-0004$30.00/0
doi:10.1016/j.jelectrocard.2004.04.004

length. While there were initially questions about the accuracy of these shorter recordings, it is now recognized that HRV from records of 2- to 5-minute duration can be used to accurately assess cardiac autonomic activity (14). Furthermore, there has been interest in using ultra-short recordings from the standard 12-lead electrocardiogram (ECG) to capture time domain measures of HRV. These records are ten seconds in duration and much easier to collect in clinical and epidemiologic settings (8,11,12).

However, the ability to make full use of HRV measures derived from both short and ultra-short term recordings is limited by the sparse data on the short-term repeatability of these measures in healthy populations (15–19). To determine the repeatability of common HRV measures derived from 10-second, 2-minute, and 6-minute recordings, we conducted a repeatability study among 63 apparently healthy, middle-aged men and women.

Materials and Methods

Data Collection and Processing

We recruited 63 participants aged 45 to 64 years from the Chapel Hill, North Carolina area. We made efforts to match the age, gender, and racial characteristics of the volunteer group to that of the Atherosclerosis Risk in Communities (ARIC) study cohort at baseline (20), and excluded volunteers if they were or may have been pregnant; if they had diagnosed diabetes, congestive heart failure, or acute or chronic renal disease; if they had a pacemaker; or if they were currently taking type 1A antiarrhythmics (quinidine, procainamide, disopyramide, or moricizine). All participants provided written informed consent, and the study was approved by the University of North Carolina at Chapel Hill Institutional Review Board.

We examined participants in the General Clinical Research Center at the University of North Carolina Hospitals between July and October 2001. Participants attended 2 visits, each approximately 60 minutes long, separated by 1 to 2 weeks. We asked participants to avoid caffeine, eating, heavy physical activity, smoking, and alcohol intake for the 10 hours preceding each clinic visit. We identified individuals who had not been fasting for 10 hours or who had smoked in the morning, but did not exclude protocol violators from participating in the study.

Four trained and certified technicians examined participants following a standardized protocol. Conditions were similar to those in the ARIC study (20–22). All examinations took place between 7:30 and 11:30 am in examination rooms maintained at a comfortable temperature with limited distractions. We described experimental procedures to participants in advance, and asked them to lie quietly before and during each recording.

After participants rested for 15 minutes in the supine position, we recorded two 10-second, standard twelve-lead resting ECGs using Kendall Q-Trace 5400 Ag/AgCl electrodes (Ludlow Co., Chicopee, MA) positioned using the E-V6 Halfpoint method (23). We then obtained one 10-second, 12-lead ECG with electrodes in the Mason-Likar position (24), and two 3-lead, 6-minute continuous recordings of R-R intervals, using the V1, V6, and RL electrodes. During the 6-minute recordings, we automatically measured blood pressure every minute using the Critikon Dinamap 8100T vital signs monitor (Critikon, Tampa FL).

We used the MAC PC Personal Cardiograph (Marquette Electronics, Inc., Jupiter, FL) to record digitized twelve-lead ECGs, which were sent via modem to the Epidemiological Cardiology Research (EPICARE) Center (Winston-Salem, NC) where they were computer coded and processed using the most recent version of the Marquette GE program version 12SL to generate the following measures: the mean heart rate across all 12 leads (HR [beats/min]); mean R-R interval duration (RR [ms]), which was the unit-corrected inverse of the mean heart rate; the standard deviation of all normal-to-normal R-R intervals (SDNN [ms]); and the root mean square of successive differences between all normal-to-normal R-R intervals (rMSSD [ms]). Unless otherwise specified, we limited the analyses of the 10-second records to those collected with the electrodes in the standard position.

While the Marquette GE program could process the 10-second records in a completely automated manner, the program for the collection and processing of the 6-minute records required human identification of artifacts. We gave each record an identification number so the 2 trained readers could not determine which records came from the same participant. Reader 1 read all 252 records once, and reread a random subset of 126 records. Reader 2 read a random subset of 63 of the 126 records twice.

We used a dedicated desktop computer and software (Arrhythmia Research Technology (ART), Inc., Austin, TX) for continuous detection of R waves at a sampling frequency of 1,000 Hz. The R waves were stored in electronic files and sent to the Ultrasound Reading Center (Winston-Salem, NC)

for data processing. Stored R-R intervals were converted into beat-to-beat heart rate data. At the reading center, readers used a modified version of the ART software to access, extract, and apply an artifact identification and imputation system to the stored files. The readers used a filter program that graphically displays a plot of beat-to-beat heart rate over time to identify and flag any visually apparent artifacts in the heart rate record. Data files were converted to binary format and transferred to a dedicated computer with specialized software (PREDICTOR II HRVECG, Arrhythmia Research Technology, Inc., Austin, TX) for data processing and spectrum analysis. The PREDICTOR II software uses a variance preserving imputation algorithm for estimating power components in the presence of artifacts. After interpolation, heart rate data were converted back to R-R intervals for computation of time domain measures and power spectral analysis, which includes time trend analysis, fast Fourier transformation, smoothing, and computation of frequency and time domain measures.

The 2-minute records were a subset of the 6-minute records, created by truncating the 6-minute records after the first 2 minutes. This truncation occurred prior to the artifact identification step. The processing of the 2-minute records then continued as described above for the 6-minute records. Further details of the processing of the 2-minute and 6-minute records have been previously published (25,26).

The calculated time domain measures included: RR, HR, SDNN, and rMSSD. For the 2- and 6-minute records, HR was the unit-corrected inverse of the mean R-R interval duration. The frequency domain measures were generated through fast Fourier transformation of the edited R-R interval records, which produces a power spectral density curve, a plot of the frequency of the cyclic components of variation in R-R interval duration against the square of their amplitude. High frequency power (HF [ms^2]), defined as the area under the power spectral density curve from 0.15 to 0.50 Hz, is regulated by the parasympathetic system (14). Low frequency power (LF [ms^2]), defined as the area under the power spectral density curve from 0.04 to 0.15 Hz, is regulated by the parasympathetic and sympathetic systems (14). In addition, we normalized high frequency and low frequency power as follows: HFnu=100% × HF/(HF+LF) and LFnu=100% × LF/(HF+LF). Use of HFnu and LFnu minimizes the effect of changes in total power on the HF and LF values (14).

The results from 10 readings (8 two-minute readings and 2 six-minute readings, corresponding to 4 different recordings) were excluded either because of extreme noise or because the proportion of acceptable R-R intervals was less than 0.80. All remaining 2- and 6-minute records had at least 30 acceptable R-R intervals and at least 60 seconds of usable recording.

All the 10-second records had signal-to-noise ratios above a pre-specified threshold. Nine records with a Minnesota Code of 8.1 (frequent premature beats), less than five acceptable R-R intervals, or the proportion of acceptable R-R intervals less than 0.50 were excluded. According to the Minnesota Codes, no records met the criteria for second or third degree atrioventricular block, Wolf-Parkinson-White Syndrome, an artificial pacemaker, or atrial fibrillation/flutter. There were also two missing ten-second records due to temporary equipment problems.

Statistical Analysis

We used nested random effects models to estimate the between-person variance, between-visit variance, and the between- and within-reader variance (27). Details are given in Appendix A. We were primarily interested in the interclass correlation coefficient (ICC), which is the proportion of the between-person variance over the total variance (ICC = σ^2_{BP}/σ^2_T, where σ^2_{BP} is the between-person variance and σ^2_T is the total variance). The ICC can also be interpreted as the correlation between different measures on the same individual, where the measures are made at different visits and read by different readers. Other correlations can be similarly defined: between repeat readings by the same reader on the same recording; between repeat readings by different readers on the same recording. These correlation coefficients may be expressed by $R_V = (\sigma^2_T - \sigma^2_V)/\sigma^2_T$, where σ^2_V is the sum of the variance components related to the type of repeatability being considered. The ICC is a special case of R_V where $\sigma^2_V = \sigma^2_T - \sigma^2_{BP}$. We present the proportion of the variance due to each source of variation. The variance due to each source of variation can be calculated by multiplying each proportion by the total variance. The coefficient of variation for each source can be calculated by dividing the square root of the variance associated with that source (ie, the standard deviation) by the mean of the heart rate variability measure.

Because the multiple recordings at a given visit on a given person were done by the same technician, models including a term for between-technician variance were, in general, unstable. Conse-

Table 1. Means for Common Heart Rate Variability Measures by Recording Length

	Arithmetic Means (Standard Deviations)			Geometric Mean		
Measure	10-seconds	2-minutes	6-minutes	10-seconds	2-minutes	6-minutes
HR (beats/min)	60 (7)	59 (8)	59 (8)	60	59	59
RR (ms)	1016 (137)	1034 (161)	1032 (160)	1008	1023	1022
SDNN (ms)	30 (32)	46 (30)	50 (30)	23	41	44
rMSSD (ms)	35 (44)	38 (35)	39 (34)	26	32	32
HF (ms^2)	—	27×10^4 (5.92×10^6)	29×10^6 (4.48×10^7)	—	13×10^4	16×10^6
LF (ms^2)	—	57×10^4 (1.34×10^7)	64×10^6 (1.10×10^8)	—	24×10^4	32×10^6
Hfnu	—	37 (21)	36 (19)	—	31	30
Lfnu	—	63 (21)	64 (19)	—	58	60

HF, high frequency power; HFnu, normalized high frequency power; HR, mean heart rate; LF, low frequency power; LFnu, normalized low frequency power; rMSSD, root mean square of successive differences in normal-to-normal RR intervals; RR, mean R-R interval; SDNN, standard deviation of all normal-to-normal RR intervals.

quently, we present models that do not include a term for the between-technician variance. All available evidence, however, indicated that the between-technician variance was negligible.

All models were implemented using the SAS MIXED procedure in SAS Version 8 (SAS Institute, Cary, NC), with the restricted maximum likelihood method. The computation of standard errors and confidence intervals depends on the assumption of normal distributions for the heart rate variability measures. Consequently, variables with skewed distributions were log-transformed in order to normalize them. Because the ICC of the variable on the original scale may be of interest to readers, we also present the ICCs on the original scale. The confidence intervals for these estimates, however, should be interpreted cautiously in light of the violation of the normality assumption. In addition to the normality assumption, assumptions of linearity and independence among effects are also required for these models. Similar models were also fit using the mean of measures derived from two or three ten-second ECGs.

We used multivariate repeated measures mixed models to estimate correlation coefficients between heart rate variability measures from different length recordings. Details of these models are given in Appendix B.

For sensitivity analyses, the above models were rerun after excluding different groups of individuals. These exclusions were performed separately, and also in various combinations. The groups included individuals who violated the fasting or smoking components of the protocol, individuals who were taking specific groups of medications (β-antagonists, thyroid hormones, α- or β-agonists, diuretics, anticholinergics, or selective serotonin receptor inhibitors), and individuals with first degree atrioventricular blocks (as defined by a PR interval greater than 200 ms) or bundle branch blocks (as defined by a QRS interval greater than 120 ms).

Results

Of the 63 participants, 31 (49%) were female and 20 (32%) were non-white. The mean age was 52, with a range of 45 to 64 years. The mean body mass index was 27 kg/m^2. The mean fasting interval for food was 13 hours and for drink was 12 hours. Six of the participants violated either the fasting or smoking protocol at one or both visits. Participants were taking a variety of medications, including selective serotonin reuptake inhibitors (n=10), anticholinergics (n=5), diuretics (n=4), α- or β-agonists (n=4), thyroid hormones (n=3), and β-antagonists (n=1).

Arithmetic and geometric means for the heart rate variability measures are shown in Table 1. SDNN, rMSSD, HF, and LF were highly skewed and R-R interval was slightly skewed. The different magnitudes of HF and LF for the 2-minute and 6-minute records are due to a different internal scaling factor that is introduced into records of different lengths by the proprietary ART software. Statistical tests comparing the means of the first recordings from the first visit, the second recordings from the first visit revealed no differences, in accordance with the assumptions underlying the subsequent statistical models.

Because ICCs are changed by logarithmic transformation, Table 2 shows the ICC for both transformed and untransformed measures. For many purposes, it is most appropriate to use the distributions that are the most normally distributed, ie, transformed values for SDNN, rMSSD, HF, and LF and untransformed values for HR, RR, HFnu and LFnu. The log-transformed values tended to have

Table 2. Intraclass Correlation Coefficients (95% Confidence Intervals) for Common Heart Rate Variability Measures by Recording Length

	Recording Length		
Measure	10-seconds	2-minutes	6-minutes
HR	0.80 (0.72-0.88)	0.89 (0.85-0.94)	0.90 (0.85-0.94)
RR	0.85 (0.78-0.91)	0.92 (0.89-0.96)	0.93 (0.90-0.96)
SDNN	0.41 (0.22-0.59)	0.86 (0.80-0.91)	0.87 (0.81-0.92)
rMSSD	0.47 (0.30-0.63)	0.91 (0.87-0.95)	0.91 (0.87-0.95)
HF	—	0.89 (0.86-0.93)	0.85 (0.79-0.91)
LF	—	0.72 (0.65-0.80)	0.83 (0.77-0.89)
Hfnu	—	0.60 (0.48-0.73)	0.76 (0.68-0.85)
Lfnu	—	0.60 (0.48-0.73)	0.76 (0.68-0.85)
ln HR	0.82 (0.75-0.90)	0.91 (0.87-0.95)	0.92 (0.88-0.95)
ln RR	0.82 (0.75-0.90)	0.91 (0.87-0.95)	0.92 (0.88-0.95)
ln SDNN	0.46 (0.31-0.62)	0.70 (0.59-0.80)	0.73 (0.63-0.83)
ln rMSSD	0.57 (0.43-0.70)	0.82 (0.75-0.89)	0.84 (0.78-0.91)
ln HF	—	0.69 (0.59-0.79)	0.82 (0.75-0.89)
ln LF	—	0.55 (0.42-0.68)	0.78 (0.70-0.86)
ln HFnu	—	0.50 (0.35-0.64)	0.76 (0.68-0.84)
ln LFnu	—	0.68 (0.58-0.78)	0.73 (0.63-0.83)

HF, high frequency power; HFnu, normalized high frequency power; HR, mean heart rate; LF, low frequency power; LFnu, normalized low frequency power; rMSSD, root mean square of successive differences in normal-to-normal RR intervals; RR, mean R-R interval; SDNN, standard deviation of all normal-to-normal RR intervals.

larger ICCs than the untransformed values for the 10-second records, and smaller ICCs for the 2-minute and 6-minute records.

While the ICCs for HR and RR were greater than 0.80 for the 10-second records, the time domain measures SDNN and rMSSD had substantially lower ICCs (Table 3). For the 2-minute and 6-minute records, the time domain measures had ICCs greater than or equal to 0.70, with little difference between 2-minute and 6-minute records. In contrast, there was substantial improvement in the repeatability of low frequency power from using 6-minute instead of 2-minute records.

Because of the comparatively low ICC values for some of measures from the 10-second records, we investigated the effect of using the average values from 2 or 3 ten-second recordings, as is frequently done in epidemiologic studies for blood pressure. The ICCs for the time domain measures were substantially improved (Table 3). Using 3 recordings increased the ICC for log-transformed SDNN from 0.46 to 0.65 and for log-transformed rMSSD from 0.57 to 0.74.

For the 2-minute and 6-minute records, the measurement error was partitioned into several components (Table 4). Note that the first line in the Table shows the proportion of the total variance that is due to the between-person component,

Table 3. Intraclass Correlation Coefficients (95% Confidence Intervals) for Four Common Heart Rate Variability Measures by Number of Consecutive Ten-second Recordings

	Number of 10-second Recordings		
Measure	One	Two	Three*
HR	0.80 (0.72-0.88)	0.83 (0.76-0.91)	0.83 (0.75-0.91)
RR	0.85 (0.78-0.91)	0.88 (0.82-0.94)	0.88 (0.82-0.94)
SDNN	0.41 (0.22-0.59)	0.46 (0.26-0.66)	0.78 (0.68-0.88)
rMSSD	0.47 (0.30-0.63)	0.53 (0.35-0.71)	0.82 (0.73-0.90)
ln HR	0.82 (0.75-0.90)	0.86 (0.79-0.92)	0.86 (0.79-0.92)
ln RR	0.82 (0.75-0.90)	0.86 (0.79-0.92)	0.86 (0.79-0.92)
ln SDNN	0.46 (0.31-0.62)	0.58 (0.41-0.75)	0.65 (0.50-0.80)
ln rMSSD	0.57 (0.43-0.70)	0.66 (0.51-0.80)	0.74 (0.62-0.85)

HR, mean heart rate; MSSD, root mean square of successive differences in normal-to-normal RR intervals; RR, mean R-R interval; SDNN, standard deviation of all normal-to-normal RR intervals.

* Includes the third recording with the electrodes in the Mason-Likar position.

Table 4. Components of Measurement Error for Common Heart Rate Variability Measures by Recording Length

Record Duration	Source of Variation	Proportion of Variance due to Each Component					
		HR	RR	SDNN	rMSSD	ln HF	ln LF
2-minutes	Between-person	0.8938	0.9234	0.8565	0.9091	0.6918	0.5511
	Between-visit	0.0802	0.0563	0.0031	0.0188	0.0267	0.0390
	Within-visit	0.0260	0.0203	0.1398	0.0717	0.2802	0.4087
	Between-reader	0.0000	0.0000	0.0004	0.0004	0.0003	0.0002
	Within-reader	0.0000	0.0000	0.0001	0.0001	0.0010	0.0009
	Total variance	61.44	26224.82	900.44	1210.32	1.27	1.50
6-minutes	Between-person	0.9002	0.9314	0.8662	0.9083	0.8114	0.7861
	Between-visit	0.0855	0.0584	0.0765	0.0670	0.0795	0.0627
	Within-visit	0.0143	0.0102	0.0554	0.0161	0.0772	0.1372
	Between-reader	0.0000	0.0000	0.0017	0.0080	0.0294	0.0083
	Within-reader	0.0000	0.0000	0.0003	0.0006	0.0025	0.0057
	Total variance	61.61	25847.80	931.95	1179.24	1.18	1.21

HR, mean heart rate; ln HF, natural log of the high frequency power; ln LF, natural log of the low frequency power; rMSSD, root mean square of successive differences in normal-to-normal RR intervals; RR, mean R-R interval; SDNN, standard deviation of all normal-to-normal RR intervals. The variances and the coefficients of variation due to each source of variation can be easily calculated from the information presented here. The variance due to each source of variation is calculated by multiplying each proportion by the total variance. The coefficient of variation is calculated by dividing the square root of the variance associated with that source (ie, the standard deviation) by the mean of the heart rate variability measure (given in Table 1).

which is the ICC. For many of the measures, the between-visit variance component is smaller than the within-visit variance component, which suggests that most of the within-person variance is due to the variability between different recordings. The added variability that comes from comparing 2 recordings from different visits versus comparing two recordings from the same visit is often small. For all of the HRV measures, the between- and within-reader components were quite small.

The correlations between the same measure from different length recordings were generally quite high, with slightly lower correlations seen comparing ten-second SDNN to two-minute, and six-minute SDNN, and for measures involving low frequency power (LF, HFnu, and LFnu) (Table 5).

In general our findings were not appreciably changed by excluding individuals who violated the fasting or smoking components of the protocol, individuals who were taking specific groups of medications, or individuals with bundle branch blocks. Excluding 9 individuals with first degree atrioventricular blocks, however, decreased the repeatability of most of the measures.

Discussion

To our knowledge, this is the first study to determine the repeatability of heart rate variability measures from the standard 10-second, 12-lead

Table 5. Correlation Coefficients (95% Confidence Intervals) for Common Heart Rate Variability Measures from Different Recording Lengths

Measure	10-second & 2-minutes	10-second & 6-minutes	2-minutes & 6-minutes
HR	0.94 (0.91-0.96)	0.94 (0.91-0.96)	0.99 (0.98-0.99)
RR	0.95 (0.93-0.97)	0.95 (0.93-0.97)	0.99 (0.98-0.99)
SDNN	0.72 (0.63-0.80)	0.76 (0.68-0.82)	0.92 (0.89-0.95)
rMSSD	0.80 (0.73-0.85)	0.82 (0.75-0.86)	0.96 (0.95-0.98)
HFnu	—	—	0.77 (0.70-0.83)
LFnu	—	—	0.77 (0.70-0.83)
ln HR	0.94 (0.92-0.96)	0.95 (0.92-0.96)	0.99 (0.98-0.99)
ln RR	0.94 (0.92-0.96)	0.95 (0.92-0.96)	0.99 (0.98-0.99)
ln SDNN	0.61 (0.50-0.70)	0.64 (0.54-0.73)	0.87 (0.83-0.91)
ln rMSSD	0.76 (0.69-0.83)	0.79 (0.72-0.84)	0.94 (0.92-0.96)
ln HF	—	—	0.80 (0.73-0.85)
ln LF	—	—	0.73 (0.65-0.80)

HF, high frequency power; HFnu, normalized high frequency power; HR, mean heart rate; LF, low frequency power; LFnu, normalized low frequency power; rMSSD, root mean square of successive differences in normal-to-normal RR intervals; RR, mean R-R interval; SDNN, standard deviation of all normal-to-normal RR intervals.

ECG. While we found low repeatability, as assessed by the intraclass correlation coefficient, taking the mean values from 3 ten-second recordings from the same recording session resulted in repeatability that compares well with that from longer records. We found that the repeatability of heart rate variability measures derived from records 2- and 6-minutes in duration was quite high, with the possible exception of low frequency power from two-minute records. Thus, records of 2-minute duration appear sufficient to obtain estimates of time domain measures and high frequency power that are repeatable and highly correlated with their 6-minute counterparts.

There has been only one previous study that investigated the measurement error of heart rate variability from ultra-short recordings (16). Our intraclass correlation coefficients for 10-second records are remarkably similar to those found by Dekker for 20-second records (16). Information about the measurement error of short-term (5-minute) heart rate variability recordings is also sparse. Previous studies have attempted to study the measurement error of short-term heart rate variability recordings using different measurement protocols and different statistics. Some studies were restricted to special populations, such as patients undergoing knee replacement (28) or patients with chronic heart failure (17,29), or were measured using very different equipment (30–32) or with controlled breathing (33–35), and thus are not strictly pertinent to this study, in which heart rate variability was measured in a presumably healthy group of middle aged adults under conditions of spontaneous breathing. Three of the 4 studies performed on healthy adults with spontaneous breathing were quite small, with samples sizes between 14 and 18 (15,16,19), while the fourth study included 70 healthy, adult residents of Israeli kibbutzim (18). In general, our intraclass correlation coefficients for 6-minute records tend to be higher than those for 5-minute records as reported by Pitzalis (19) and Dekker (16), and comparable to those found by Ahmed (15) and Sinnreich (18). A recent study computed the repeatability of 3-minute measures using 36 older adults, with a mix of healthy and diabetic subjects, and found lower repeatability than that of our 2-minute measures (36).

In general, the ICC values that we found for 2- and 6-minute measures and for the mean of three 10-second measures compares favorably with those found for other commonly used measurements in the ARIC study. For example, unpublished results from 190 ARIC participants with repeat measurements 1 to 2 weeks apart gave ICCs of 0.75 for systolic blood pressure and 0.62 for diastolic blood pressure when using the average of the second and third blood pressure measurements.

The correlation coefficients that we report between 10-second and 6-minute measures are very similar to the Spearman correlation coefficients that Dekker computed between 20-second and 5-minute records (0.96 for HR, 0.48 for SDNN, and 0.73 for rMSSD) (16). Our correlations between 2- and 6-minute measures are similar to those that Marks reported for 2.5- and 5-minute records under controlled breathing (0.99 for HR, 0.98-0.99 for RR, 0.99 for SDNN, 0.53-0.67 for LF, and 0.71-0.72 for HF) (34).

This study also partitioned the measurement variance into several different components, thus shedding light on potentially important sources of error. It must be noted that this study used very standardized collection and processing procedures which minimize measurement error. We were able to recruit a middle-aged study population that was approximately half female and half minority, which should facilitate extrapolation to other population groups. However, about half of our sample had mean heart rates less than 60 beats/min, which may limit the generalizability of our findings. Furthermore, caution should be used when extrapolating these results to hospitalized or patient populations, or to clinical settings in which similar attention is not given to minimizing measurement error.

Because of the design of the study, it was impossible to fully separate the inter-technician variance component, although available evidence suggests that this variance component was quite negligible. The durations that we studied, 10-seconds, 2-minutes, and 6-minutes, correspond to the duration of extant records from the ARIC study, and were chosen in order to permit application of these estimates to the ARIC study. Due to resource constraints, we were unable to examine 5-minute records, which is the standard currently recommended by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (14). Based on the data we have presented for 2-minute and 6-minute records; however, it should be possible to approximate values for 5-minute records.

The information we present on the repeatability of heart rate variability measures derived from short and ultra-short term records should be of value to researchers planning future studies of heart rate variability. We recommend using records at least 5-minutes in duration, when possible, in accordance with the previous guidelines (14). In

large scale epidemiologic studies where longer records are unavailable, it may be possible to use time domain measures derived from a single 10-second recording. However, researchers using 10-second records should consider taking the mean of several recordings. If only a single recording is available, researchers should strongly consider statistical methods to adjust for the considerable measurement error (37–39).

Acknowledgment

The authors thank Drs. Dana Loomis and Wayne Rosamond for their review of and comments on drafts of the manuscript.

References

1. Kleiger RE, Miller JP, Bigger JT Jr, et al: Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. Am J Cardiol 59:256, 1987
2. Vaishnav S, Stevenson R, Marchant B, et al: Relation between heart rate variability early after acute myocardial infarction and long-term mortality. Am J Cardiol 73:653, 1994
3. Lombardi F, Sandrone G, Pernpruner S, et al: Heart rate variability as an index of sympathovagal interaction after acute myocardial infarction. Am J Cardiol 60:1239, 1987
4. Malik M, Farrell T, Camm AJ: Circadian rhythm of heart rate variability after acute myocardial infarction and its influence on the prognostic value of heart rate variability. Am J Cardiol 66:1049, 1990
5. Bigger JT, Jr, Fleiss JL, Steinman RC, et al: Frequency domain measures of heart period variability and mortality after myocardial infarction. Circulation 85: 164, 1992
6. Liao D, Cai J, Rosamond WD, et al: Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study: the ARIC Study. Am J Epidemiol 145:696, 1997
7. Dekker JM, Crow RS, Folsom AR, et al: 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, 2000
8. Tibblin G, Eriksson C-G, Bjurö T, et al: Heart rate and heart rate variability. A risk factor for the development of ischemic heart disease (IHD) in the "Men Born in 1913 Study"-A ten year follow-up. IRCS Medical Science. Cardiovascular System: Social and Occupational Medicine 3:95, 1975
9. Liao D, Cai J, Barnes RW, et al: Association of cardiac autonomic function and the development of hypertension: the ARIC Study. Am J Hypertens 9:1147, 1996
10. Singh JP, Larson MG, Tsuji H, et al: Reduced heart rate variability and new-onset hypertension: insights into pathogenesis of hypertension: the Framingham Heart Study. Hypertension 32:293, 1998
11. Dekker JM, Schouten EG, Klootwijk P, et al: Heart rate variability from short electrocardiographic recordings predicts mortality from all causes in middle-aged and elderly men: The Zutphen Study. Am J Epidemiol 145:899, 1997
12. de Bruyne MC, Kors JA, Hoes AW, et al: Both decreased and increased heart rate variability on the standard 10-second electrocardiogram predict cardiac mortality in the elderly: the Rotterdam Study. Am J Epidemiol 150:1282, 1999
13. Tsuji H, Venditti FJ Jr, Manders ES, et al: Reduced heart rate variability and mortality risk in an elderly cohort: the Framingham Heart Study. Circulation 90:878, 1994
14. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93:1043, 1996
15. Ahmed MW, Kadish AH, Parker MA, et al: Effect of physiologic and pharmacologic adrenergic stimulation on heart rate variability. J Am Coll Cardiol 24:1082, 1994
16. Dekker JM, de Vries EL, Lengton RR, et al: Reproducibility and comparability of short- and long-term heart rate variability measures in healthy young men. Ann Noninvasive Electrocardiol 1:287, 1996
17. Ponikowski P, Piepoli M, Amadi AA, et al: Reproducibility of heart rate variability measures in patients with chronic heart failure. Clin Sci 91:391, 1996
18. Sinnreich R, Kark JD, Friedlander Y, et al: Five minute recordings of heart rate variability for population studies: Repeatability and age-sex characteristics. Heart 80:156, 1998
19. Pitzalis MV, Mastropasqua F, Massari F, et al: Short- and long-term reproducibility of time and frequency domain heart rate variability measurements in normal subjects. Cardiovasc Res 32:226, 1996
20. ARIC Investigators: The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 129:687, 1989
21. National Heart Lung and Blood Institute: The ARIC manuals of operation. Manual 16. Heart rate variability data collection manual, Visit 4. University of North Carolina at Chapel Hill, ARIC Coordinating Center, Chapel Hill, NC, 1998.
22. National Heart Lung and Blood Institute: The ARIC manuals of operation. Manual 5. Electrocardiography, Visit 4. University of North Carolina at Chapel Hill, ARIC Coordinating Center, Chapel Hill, NC, 1997
23. Rautaharju PM, Wolf HK, Eifler WJ, et al: A simple procedure for positioning precordial ECG and VCG

electrodes using an electrode locator. J Electrocardiol 9:35, 1976
24. Mason RE, Likar I: A new system of multiple-lead exercise electrocardiography. Am Heart J 71:196, 1966
25. Liao D, Creason J, Shy C, et al: Daily variation of particulate air pollution and poor cardiac autonomic control in the elderly. Environ Health Perspect 107: 521, 1999
26. Liao D, Barnes RW, Chambless LE, et al: A computer algorithm to impute interrupted heart rate data for the spectral analysis of heart rate variability-The ARIC Study. Comput Biomed Res 29:140, 1996
27. Littell RC, Milliken GA, Stroup WW, et al: SAS System for Mixed Models. SAS Institute, Cary, NC, 1996
28. Freed LA, Stein KM, Gordon M, et al: Reproducibility of power spectral measures of heart rate variability obtained from short-term sampling periods. Am J Cardiol 74:972, 1994
29. Piepoli M, Radaelli A, Ponikowski P, et al: Reproducibility of heart rate variability indices during exercise stress testing and inotrope infusion in chronic heart failure patients. Clin Sci 91:87, 1996
30. Cloarec-Blanchard L, Funck-Brentano C, Lipski M, et al: Repeatability of spectral components of short-term blood pressure and heart rate variability during acute sympathetic activation in healthy young male subjects. Clin Sci 93:21, 1997
31. Dimier-David L, Billon N, Costagliola D, et al: Reproducibility of non-invasive measurement and of short-term variability of blood pressure and heart rate in healthy volunteers. Br J Clin Pharmacol 38:109, 1994
32. Parati G, Omboni S, Villani A, et al: Reproducibility of beat-by-beat blood pressure and heart rate variability. Blood Press Monit 6:217, 2001
33. Töyry J, Mäntysaari M, Hartikainen J, et al: Day-to-day variability of cardiac autonomic regulation parameters in normal subjects. Clin Physiol 15:39, 1995
34. Marks BL, Lightfoot JT: Reproducibility of resting heart rate variability with short sampling periods. Can J Appl Physiol 24:337, 1999
35. Amara CE, Wolfe LA: Reliability of noninvasive methods to measure cardiac autonomic function. Can J Appl Physiol 23:396, 1998
36. Gerritsen J, TenVoorde BJ, Dekker JM, et al: Measures of cardiovascular autonomic nervous function: Agreement, reproducibility, and reference values in middle age and elderly subjects. Diabetologia 46:330, 2003
37. Chambless LE, Davis V: Analysis of associations with change in a multivariate outcome variable when baseline is subject to measurement error. Stat Med 22:1041, 2003
38. Carroll RJ, Ruppert D, Stefanski LA: Measurement Error in Nonlinear Models. Chapman & Hall, New York, 1995
39. Fuller WA: Measurement Error Models. John Wiley & Sons, New York, 1987

Appendix A

We applied the following random effects model to the two-minute and six-minute heart rate variability measures (27):

$$Y_{ijklm} = \mu + PERSON_i + VISIT_j(PERSON_i) + RECORD_k(PERSON_i, VISIT_j) + READER_l(PERSON_i, VISIT_j, RECORD_k) + ERROR_{ijklm}$$

where Y_{ijklm} = heart rate variability measures and μ is the intercept. The terms in the model are assumed independent. From this model one can write the total variance as:

$$Var(Y_{ijklm}) = \sigma^2_{TOTAL(T)} = \sigma^2_{BP} + \sigma^2_{BV} + \sigma^2_{WV} + \sigma^2_{BR} + \sigma^2_{WR}$$

σ^2_{BP} is the between-person variation and σ^2_{BV} is the between-visit variation. σ^2_{WV} is the within-visit variation, which comes from different recordings taken at the same visit on the same person by the same technician. σ^2_{BR} is the between-reader variation and σ^2_{WR} is the within-reader variation, from the error term. Thus $\sigma^2_T - \sigma^2_{BP}$ is the total measurement variance, where σ^2_T is the total of all the variance components.

Because the reading process is completely automated for the ten-second records, a simpler model was used: $Y_{ijk} = \mu + PERSON_i + VISIT_j(PERSON_i) + ERROR_{ijk}$. The total variance for the ten-second records is thus: $Var(Y_{ijk}) = \sigma^2_{TOTAL(T)} = \sigma^2_{BP} + \sigma^2_{BV} + \sigma^2_{WV}$.

Appendix B

A multivariate repeated measures model was used to compute correlations between different heart rate variability measures, assuming distinct means and variances for measurements from different length recordings. To compute the correlation coefficient between a heart rate variability measure from a 10-second recording and the same measure from a 6-minute recording, we used a vector of 8 measurements, in the following order: first 10-second recording from the first visit, second 10-second recording from the first visit, first 10-second recording from the second visit, second 10-second recording from the second visit, first 6-minute recording from the first visit, second 6-minute recording from the first visit, first 6-minute record-

ing from the second visit, and the second 6-minute recording from the second visit. This vector was assumed to have a covariance matrix Σ, with the following structure:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & & & & & \\ \sigma_{21} & \sigma_1^2 & & & & & & \\ \sigma_{31} & \sigma_{31} & \sigma_2^2 & & & & & \\ \sigma_{31} & \sigma_{31} & \sigma_{43} & \sigma_2^2 & & & & \\ \sigma_{51} & \sigma_{51} & \sigma_{71} & \sigma_{71} & \sigma_1^2 & & & \\ \sigma_{51} & \sigma_{51} & \sigma_{71} & \sigma_{71} & \sigma_{21} & \sigma_1^2 & & \\ \sigma_{71} & \sigma_{71} & \sigma_{73} & \sigma_{73} & \sigma_{31} & \sigma_{31} & \sigma_2^2 & \\ \sigma_{71} & \sigma_{71} & \sigma_{73} & \sigma_{73} & \sigma_{31} & \sigma_{31} & \sigma_{43} & \sigma_2^2 \end{pmatrix}$$

This covariance matrix has 8 parameters, and contains the following assumptions: the variance of all the 10-second measures is the same, and equals σ_1^2; the variance of all the 6-minute measures is the same, and equals σ_2^2; the covariance between the 2 different 10-second measures at a given visit is σ_{21}; the covariance between the 2 different 6-minute measures at a given visit is σ_{43}; the covariance between any given 10-second measure and any given 6-minute measure at the same visit is σ_{31}; the covariance between any given 10-second measure at one visit and any given 10-second measure at the other visit is σ_{51}; the covariance between any given 6-minute measure at one visit and any given 6-minute measure at the other visit is σ_{73}; and the covariance between any given 10-second measure at 1 visit and any given 6-minute measure at the other visit is σ_{71}. Analogous models were used to compare measurements from 10-second and 2-minute recordings, and from 2-minute and 6-minute recordings.

An important assumption for the above model is that the means of the measures do not vary by visit and by order within visit. In order to test this assumption, repeated measures mixed models were fit with a fixed effect term for the visit, recording combination (with 4 levels). The F test for the fixed effect term was used to test for inequality of means. In all cases, the *P* value was extremely high, and did not support the hypothesis that any of the means are unequal.