Reliability of Respiratory Sinus Arrhythmia in R-R intervals, in 14, 20, and 26 week old infants

John E. Richards
Department of Psychology
University of South Carolina
Columbia, SC 29208

Running head: RSA Reliability

Abstract

Measures of respiratory sinus arrhythmia (RSA) in R-R intervals were calculated from a five minute recording of 3 to 6 month old full-term infants. The reliability of the measures was estimated with Cronbach's α for sampling durations of 5, 15, 25, or 60 seconds, for single or multiple samples from each of the five baseline periods. Reliability of RSA was good (> .8) for samples of 25 seconds or greater, but decreased with smaller sampling durations. The sensitivity of the measures in detecting an age effect between 14, 20, and 26 week old infants correlated positively with the reliability of the measure. This study suggests that reliability of RSA measures in young infants is acceptable.

Reliability of Respiratory Sinus Arrhythmia in R-R intervals, in 14, 20, and 26 week old infants

The variability of R-R intervals (RRV) in the electrocardiogram of young infants is an important topic (R-R interval and heart period are used interchangeably). Many methods have been proposed for measurement of RRV (e.g., van Geign, Jongsma, deHaan, & Eskes, 1980; Harper, Scalabassi, & Estrin, 1976; Yeh, Forsythe, & Hon, 1973). Of these, most distinguish between RRV occurring at different frequencies. A common distinction made is between long and short-term variability (Yeh et al., 1973). Use of spectral analysis techniques (Akselrod et al., 1985; Harper et al., 1976; Womack, 1971) allows the distinction of variability at specific frequencies. Blood pressure cardiovascular rhythms cause low frequency RRV, and are affected by sympathetic and parasympathetic activity on the heart (Akselrod et al., 1981, 1985). Vagal parasympathetic activity causes RRV occurring at the frequency of respiration, respiratory sinus arrhythmia (RSA) (Akselrod et al., 1985; Anrep, Pascual, & Rossler, 1935; Katona & Jih, 1975; Lopes & Palmer, 1976; Porges, McCabe, & Yongue, 1982). Changes in the parasympathetic activity affecting R-R interval length partially cause developmental changes in RSA level (Harper et al., 1976, 1978; Egbert & Katona, 1980; Watanabe, Iwase, & Hara, 1973). RSA has been shown to be uniquely related to behavioral and pathological status in newborns and young infants, such as neonatal respiratory distress (Divon et al., 1986; Kero, 1973; Rother et al., 1987), sudden infant death syndrome (e.g., Harper et al., 1982), and socioemotional and cognitive processes (e.g., Fox, 1989; Richards, 1988; Richards & Casey, 1992).

One measurement characteristic of R-R interval variability methods that is important is reliability. Reliability is the extent to which an individual's observed score on a variable reflects the individual's true score. High levels of reliability imply that the observed variable is concomitant with the true score. Measures of reliability (e.g.,

Cronbach's α, Cronbach, 1951) range between 1.0 and 0, with 1.0 being perfectly reliable and 0 being completely unreliable. The reliability coefficient can be interpreted as the correlation between the observed, measurable score, and the underlying true score. Theoretically, reliability is important because many research studies use R-R interval variability as a measure of a stable individual score. The distinction between different groups (e.g., pre- and full-term infants), pathological and normal outcomes (e.g., CNS functioning, SIDS), or individual differences within a group (e.g., attention or temperament), demands that individuals retain their relative standing on the score on different testing occasions.

An issue concerning heart rate variability measurement is how many observations of a single individual are needed to obtain reliable scores, or how long one must assess heart rate in order to obtain reliable scores. At least one study investigated this aspect of reliability (Detwiler, Jarisch, & Caritis, 1980). Computer-generated time series representing "fetal heart rate" were analyzed in intervals from 20 to 120 seconds, with heart rates from 90 to 210 bpm. As expected from statistical sampling properties, the longer the sampling interval, and the number of beats sampled, the more accurately the mean value of the generated time series was produced. For example, the 95% confidence interval for measures of high frequency variability recorded for 30 seconds with heart rate at 150 bpm was about 32% of the variability measure. It was estimated that five minutes would have been needed to get a 95% confidence interval of 10%. Such measures, often taken in patients in short intervals compared to these, may be inaccurate measures of the true underlying score. This study would suggest that recording intervals of 1 minute or less have low reliability in estimating underlying true scores.

The present study tested the reliability of some methods of quantifying respiratory sinus arrhythmia with a single five-minute recording session (Note 1). The

electrocardiogram was recorded in a five minute baseline recording in 14, 20, and 26 week old infants. RSA measures were calculated from these recordings. Two quantification parameters were tested in this study. First, the R-R interval variability was quantified at several frequency bands, all of which are thought to include variability occurring at the respiratory frequency (i.e., RSA). Three measures were taken that involve a broad band of frequencies. Second, different sampling algorithms were examined. The algorithms differed in the duration of the individual measures (5 to 60 seconds), or the number of samples used from each minute (1, 2, or 3). The measure of reliability used was Cronbach's α (Cronbach, 1951). Cronbach's α is defined for variables that consist of linear composites of other variables, and is computed from the average correlation between the composite variables. Thus, for the measures of various durations, the "composite" variable consisted of the variable constructed by adding up the estimates from the five 1-minute epochs.

Methods

Subjects

The subjects were 101 full term infants (gestational age greater than 38 weeks, birthweight greater than 2500 grams) sampled cross-sectionally at 14 (N = 33, *M* = 100.5 days, *SD* = 4.82), 20 (N = 34, *M* = 141.7 days, *SD* = 2.62), or 26 (N = 34, *M* = 184.5 days, *SD* = 3.83) weeks postnatal age. The infants had no acute or chronic pre- or perinatal medical complications, and were in good health at the time of recording.

Procedure

The electrocardiogram (ECG) and respiration were recorded for five minutes while the infant sat on its parent's lap on a couch during an awake and alert state. The ECG was recorded by placing Ag-AgCl electrodes on the infant's chest, and was digitized at 1000 Hz (each msec) with a microcomputer. The peak of the QRS complex was identified in the ECG, a computer algorithm identified the peak of the R wave, and

R-R interval was defined as the duration between successive R-waves in the ECG. Artifact correction was done by first identifying suspect R-R intervals using the Cheung (1981) and Berntson, Quigley, Jang, & Boysen (1990) algorithms. Visual inspection of the ECG recording and each identified beat was done on computer displays. The suspect beats were eliminated if they did not occur on an indentifiable R-wave peak in the ECG, or adjusted to fit a R-wave peak. Respiration was measured with a pneumatic chest cuff, and was digitized on-line at 50 Hz (each 20 msec). The peak and trough of the digitized respiration recording, representing inspiration and expiration, were identified by computer algorithms. Visual inspection of the respiration recording and each identified breath was done on computer displays to eliminate artifacts.

Quantification of variability

The measures of variability were computed on R-R intervals. The values of each beat were calculated on a 0.1-s by 0.1-s basis by assigning to 100 msec bins the value of the interval weighted by the proportion of time any beat occupied that interval. Each variable was calculated separately on the five one-minute periods of the recording. For each variable, when possible, the variable was calculated over the entire 1 minute, 25 second, 15 second, or 5 second intervals. For the results analysis, either a single sample or multiple samples were taken from each of the five baseline minutes. The multiple samples were taken from the first and second halves of the minute (for ten 25-s samples), from the minutes evenly divided into thirds (fifteen 15-s or 5-s samples), or from the middle portion of each minute (five 25-s, 15-s, or 5-s samples).

Three methods (spectral analysis, band-pass filter, and peak-to-trough filter) were used to obtain variability measures at a specific frequency band (0.333-1.25 Hz) (Note 1). This band is the "high frequency" band of three frequencies typically extracted from R-R interval variability in infants (Divon et al., 1986; Schechtman, Kluge, & Harper., 1988). The high frequency variability is thought to be respiratory sinus arrhythmia. The

three frequency methods were also used to obtain an estimate of variability occurring at the same frequency of respiration. Respiration frequency was defined as the frequency at which the largest peak in the power spectrum of respiration occurred. The peak was found for each minute of the baseline. Three other methods were used that contain variability across a broad band of frequencies. The filtering methods were first applied to the entire recording minute before extracting the shorter segments.

Spectral analysis (SPA; Askelrod et al., 1985; Harper et al., 1976; Porges et al., 1982; Richards, 1986; Womack, 1971). Power spectral analysis was used to quantify respiratory sinus arrhythmia (RSA) by estimating the power of the R-R intervals and summing the power estimates at specific frequencies. The periodogram was computed with the Fast Fourier Transform from values assigned to the first 512 0.1-s intervals of each of the minutes (cosine tapered), giving a frequency resolution of 0.01953 Hz. A modified Daniell smoothing algorithm was applied to the periodogram to obtain the power spectrum. RSA was defined as the natural logarithm of the sum of the power estimates across the frequency intervals for the fixed frequency band (0.333-1.25 Hz). For the respiration frequency, the value was defined as the power summed over 0.1953 Hz (11.71 breaths per minute) and centered at the respiration frequency. Only the 1 minute sampling interval was used, since shorter intervals produce spurious results with spectral analysis. The natural logarithm of the spectral power was used.

Band-pass filter (FILTER; Porges, 1986a, 1986b). A method was used with moving average band-pass filters with predefined frequency characteristics, and the standard deviation of the filtered data was the estimate of R-R interval variability. The filtering was done by separate passes of low-pass moving average filters, and a subtracting method to get the filtered time series. Values for the R-R intervals were assigned to equidistant 20 msec bins (50 Hz sampling rate). A symmetric trigonometric moving average low-pass filter was applied to the equidistant R-R time series . The

resultant filtered data was subtracted from the original series, resulting in a high-pass series. The same trigonometric function was applied to the high-pass series with weights for a higher frequency. This resulted in a band-pass filtered series with low and high frequencies defined by the filters in the first and second pass. The frequencies for high and respiration frequencies were obtained by using different lengths of the trigonometric coefficients for the moving average filter. For the respiration frequency, the value was defined by filters spanning 0.333 Hz and centered at the respiration frequency. The natural logarithm of the standard deviation of the filtered data was the R-R interval variability measure.

Peak-to-trough filter (PTF; Schechtman et al., 1988). The third method was a time-domain method that identifies peaks and troughs in the R-R intervals, and estimates the frequency of peak-to-peak intervals by the duration of the interval, and estimates R-R interval variability as the peak-to-trough difference. For the fixed frequency, the frequency was defined by the duration of the peak-to-peak interval in the R-R record. The estimate for the respiration frequency was formed by the durations of the peaks that fell into the range spanning 0.333 Hz and centered at the respiration frequency. There were two differences between the method defined by Schechtman et al. (1988) and the current study: 1) the mean of the difference is reported rather than the median, since many of the short sampling durations only had 2 or 3 data points, making the median as unstable as the mean. The reliability analyses came out the same for the mean and median for those intervals where both the mean and median were available. 2) several subjects did not contribute a measure in a specific interval, especially for the analyses using 5, 15, or 25 s sample duration. The missing data occurred because in the selected epoch for analysis (e.g., 5, 15, or 25 s), a peak-to-peak sequence of R-R intervals matching the appropriate frequency did not occur.

Broadband filters Three measures were "broadband" filters, that include a broad range of frequencies of interest. Standard deviation of R-R intervals (HPSD): The standard deviation of the R-R intervals, heart period standard deviation, includes variation at all frequencies. Mean squared successive differences (MSSD): The mean square successive differences measure is computed as the standard deviation of subtracting successive intervals. Successive differences indices are filters that pass all frequencies above a certain frequency (Wastell, 1981). Given fixed heart rate of 120 bpm (R-R interval = 500 msec), 0.5 Hz frequency variability would be passed at 50% level, and greater frequencies would be passed at greater levels. For 150 bpm, the 50% level would be 0.62 Hz. Given the average heart rate in the present study, MSSD passes variability at 50% level at frequencies greater than approximately 0.55 Hz. Thus, the MSSD acted as a high-pass filter, passing frequencies greater than 0.55 Hz. Band-pass filter (BPF): A band-pass filter that allows variability in a precisely defined frequency interval (0.30 to 1.30 Hz) was computed using methods and coefficients of Porges (i.e., "V"; Porges, 1985, 1986a, 1986b). The BPF and FILTER measures were similarly derived (e.g., low-pass step, subtraction for high-pass series, subsequent filter for elimination of high frequencies, natural logarithm of the standard deviation of the filtered series was the R-R interval variability measure). There were several differences between the BPF and FILTER measures: 1) the BPF used 200 msec (5 Hz) rather than 20 msec (50 Hz) bins for R-R intervals, and correspondingly different coefficient lengths; 2) the filter applied for the elimination of high frequencies in the BPF was a band-pass filter, rather than a second pass of a low-pass filter with higher frequency cutoff in FILTER; 3) the low-pass filter for BPF was based on polynomial rather than trigonometric coefficients; 4) the coefficients for the low-pass step and for the elimination of the high frequencies were based on different mathematical formulae in BPF, whereas they were based on the same trigonometric equation in FILTER; 5) the

frequency band for BPF was different than the high frequency FILTER. The transfer function for the BPF and high frequency FILTER were similar. The FILTER measure had the advantage of calculating a measure centered at the infant's respiration frequency for the baseline minute upon which the score was derived.

Results

The R-R interval variability reliabilities are in Table 1. Many of the reliability levels were above .8, indicating a good reliability. Overall, there was a decrease in reliability with decreasing sampling duration. Figure 1 shows the reliability levels as a function of sample duration for the high, respiration, and broadband frequency variability, averaged across all measures for the three frequency types. The reliability of the 1 minute sample, and the samples with multiple 25, 15, or 5 second durations, were generally above .80.

Insert Table 1 and Figure 1 here

The correlations for the scores summed over the complete five minute recording were examined within each measure for the different sampling intervals. As expected from the pattern of reliabilities, the correlations were high, but showed attenuation as the measures were from samples of smaller duration. The decreasing correlations parallel the decreasing reliability levels as a function of how long the sampling duration was. For example, the correlations for the HPSD between the 1 minute sample, and the 10:25s, 15:15s, 5:25s, 15:5s, 5:15s, and 5:5s were .967, .929, .924, .806, .782, and .632, respectively. This pattern held over the variability measures, and the level of the decrease in correlations paralleled the level of the decreasing reliabilities over the sampling intervals.

Insert Table 2 here

Several studies (e.g., Harper et al., 1976, 1978; Egbert & Katona, 1980; Katona, Frasz, & Egbert, 1980; Watanabe et al., 1973) have shown that respiratory sinus arrhythmia increases with age over the age range of the infants in this study. Therefore, separate one-way ANOVA's were done with each of the R-R interval variability

measures as dependent variables, and with age as a factor. Table 2 shows the probability levels for the age effect on R-R interval variability. Many of the variables were significantly affected by age, with variability increasing over the 14 to 26 week old period.

The sampling duration showed a some relation to the probability level of the age effect, with longer intervals being more likely to detect a significant age effect. This was probably caused by the different reliabilities. The quantification techniques for respiratory sinus arrhythmia for the longest sampling intervals (the most reliable; Table 1) were sensitive to the age effect, whereas several of the measures at the shorter duration intervals (least reliable; Table 1) were not. Figure 2 shows a plot of the p-values as a function of mean reliability level within each type of frequency. This figure illustrates the relation between reliability and p-value, with reliability levels above .80 necessary for detecting a significant age effect.

Insert Figure 2 here

Discussion

The reliability analyses imply that some measures of respiratory sinus arrhythmia (RSA) have an adequate reliability level. The reliabilities of .8 and greater imply that these measures are assessing the underlying "true score" well. On the other hand, some of the reliabilities were less than .7, meaning that the hypothetical true score was not accurately measured by the computed RSA. The duration of the sampling interval was positively related to reliability. For durations of 60 seconds, and durations of 25, 15, and 5 s when multiple observations occurred in each minute, respiratory sinus arrhythmia reliability was above .8 in almost every instance. Selecting the specific respiration frequencies did not improve RSA measurement. The reliabilities of the fixed high frequency (0.333 - 1.25 Hz) and the respiration frequency measures were substantially equivalent. Selecting the observed respiration frequency to quantify the measures did not aid in reliability level (Table 1) or detecting age effects (Table 2).

These results are similar to those found by Detwiler et al. (1980). They sampled simulated heart rate data for 20, 30, 60, 90, or 120 seconds. This corresponds to the 5:5s, 5:15s, 15:5s, and 5:25s sampling strategies in the current study, that were 25, 75, 75, and 125 seconds in total duration. The poorest reliabilities occurred for the 5:5s (25 s total) and 5:15s (75 s total) sampling patterns. However, reliability levels and sensitivity to the age effect for the high frequency respiratory sinus arrhythmia at the 15:5s and 5:25s samples did not differ from the longer durations. Sampling for 75 seconds may be satisfactory if the samples are taken in multiple, separated segments of time, particularly for the measurement of high frequency R-R interval variability. Thus, for clinical and experimental situations in which the duration of the event is short (5 to 15 seconds), such as in fetal heart rate between adjacent uterine contractions, or short latency attention episodes, a single observation may be insufficient to estimate respiratory sinus arrhythmia. Several such episodes combined together should provide a sufficiently accurate respiratory sinus arrhythmia estimate.

Overall, the duration of the sample was positively related to reliability levels and the ability to detect significant age effects. If the number of samples was held constant and the sample duration increased, then reliability levels were higher for the longer intervals (e.g., 15:5s vs 15:15s; 5:5s vs 5:15s; 5:15s and 5:25s vs 5:60s). This was not true in all cases in the present study (e.g., 5:15s to 5:25s), but such exceptions may have resulted from the specific samples taken in the present study. The sampling method itself also contributed to reliability levels. Sampling for multiple separated segments of time is a better procedure than sampling fewer times for longer intervals. For example, the 15:5s (75 s total duration) sampling method resulted in higher reliability levels than did the 5:15s (75 s total duration) or the 5:25s (125 s total duration) methods. Both the 15:15s (225 s total duration) and the 10:25s (250 s total duration) resulted in higher reliability levels than the 5:60s (300 s total duration) method, even though the

latter method resulted in a longer duration sample. The probability of detecting a significant age effect was more closely related to reliability levels than sampling duration and/or sampling method, although this relation was not exact (Figure 2).

On the basis of these results, some recommendations can be made about specific measures. For the longest sampling intervals, all the measures provided equal reliability levels. As sampling duration decreases, the moving average filtering techniques (FILTER, high or respiration frequencies; and BPF, which is Porges' V) hold up better than the others. For example, the filter techniques retained their reliability level better than did the peak-to-trough method (Schectman et al., 1988). The filtering techniques were also more sensitive to the age effect across several sampling intervals. For the peak-to-trough measure, there were intervals where a subject did not contribute a measure in a specific interval (i.e., 10:25s, 15:5s). Thus, although its reliability was similar to the other measures within the same frequency, this measure can only be used with one minute segments to get appropriate data. Selecting the specific respiration frequencies did not improve the measurement of respiratory sinus arrhythmia. It did not aid in reliability level or detecting age effects. The only significant source of variation in the range from about .3 to 1.5 Hz must have been respiratory sinus arrhythmia.

Differing reliability levels of the RSA measures would have an important effect in psychological experiments in differentially attenuating the correlation between RSA and other variables. The sensitivity of the analysis of effects on RSA of such variables as age, risk status, or individual differences, will be limited by the reliability of the RSA measures. The current study showed that RSA measures with reliabilities of .85 or better were sensitive to the age effect, whereas those with reliability below .80 were not. An absolute cutoff for reliabilities cannot be made. The absolute value of reliability that changes an effect from "significant" to "nonsignificant" depends on the strength of the relation between the true scores of the two variables, and the reliabilities of the two

variables. A reliability of .5 may be sufficient to detect a significant effect in some situations (very strong relation), whereas in other situations a reliability of .9 would be insufficient (very weak relation). One strategy to solve this problem would be to use the known reliabilities of the RSA measures (Table 1) to disattenuate correlations between RSA and other variables in order to estimate the true strength of the effect.

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 fluctuations: A quantitative probe of beat-to-beat cardiovascular control. *Science, 213,* 220-222.

Anrep, G.W., Pascual, W., & Rossler, R. (1935). Respiratory variations of the heart rate. I. The reflex mechanism of the respiratory arrhythmia. *Royal Society of London Proceedings, Series B, 119,* 191-217.

Berntson, G. G., Quigley, K.S., Jang, J.F., & Boysen, S.T. (1990). An approach to artifact identification: Application to heart period data. *Psychophysiology, 27,* 586-598.

Cheung, M.N. (1981). Detection and recovery from errors in cardiac interbeat intervals. *Psychophysiology, 18,* 341-346.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Detwiler, J.S., Jarisch, W., & Caritis, S.N. (1980). Statistical fluctuations in heart rate variability indices. *American Journal of Obstetrics and Gynecology, 136,* 243-248.

Divon, M.Y., Winkler, H., Yeh, S.Y., Platt, L.D., Langer, O., & Merkatz, I.R. (1986). Diminished respiratory sinus arrhythmia in asphyxiated term infants. American *Journal of Obstetrics and Gynecology, 155,* 1263-1266.

Egbert, J.R., & Katona, P.G. (1980). Development of autonomic heart rate control in the kitten during sleep. *American Journal of Physiology, 238,* H829-H835.

Fox, N.A. (1989). Psychophysiological correlates of emotional reactivity during the first

year of life. *Developmental Psychology, 25,* 364-372.

van Geign, H.P., Jongsma, H.W., deHaan, J., & Eskes, T.K.A.B. (1980). Analysis of heart rate and beat-to-beat variability: Interval difference index. *American Journal of Obstetrics and Gynecology, 138,* 246-252.

Harper, R.M., Hoppenbrouwers, T., Sterman, M.B., McGinty, D.J., & Hodgman, J. (1976). Polygraphic studies of normal infants during the first six months of life. I. Heart rate and variability as a function of state. *Pediatric Research, 10,* 945-951.

Harper, R.M., Leake, B., Hodgman, J.E., & Hoppenbrowers, T. (1982). Developmental patterns of heart rate and heart rate variability during sleep and waking in normal infants and infants at risk for the sudden infant death syndrome. *Sleep, 5,* 28-38.

Harper, R.M., Scalabassi, R.J., and Estrin, T. (1976). Time series analysis and sleep research. *IEEE Transactions on Automatic Control, AC-19,* 932-943.

Harper, R.M., Walter, D.O., Leake, B., Hoffman, H.J., Sieck, G.C., Sterman, M.B., Hoppenbrouwers, T., & Hodgman, J. (1978). Development of sinus arrhythmia during sleeping and waking states in normal infants. *Sleep, 1,* 33-48.

Katona, P.G., Frasz, A., & Egbert, J.R. (1980). Maturation of cardiac control in full-term and preterm infants during sleep. *Early Human Development, 4,* 145-159.

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

Kero, P. (1973). Heart rate variation in infants with the Respiratory Distress Syndrome. *Acta Paediatrica Scandanavica, 250.*

Lopes, O.U., & Palmer, J.F. (1976). Proposed respiratory 'gating' mechanism for cardiac slowing. *Nature, 264,* 454-456.

Porges, S.W. (1985). Method and apparatus for evaluating rhythm in oscillations in aperiodic physiological response systems. United States Patent No 4,510,944.

Porges, S.W. (1986a). V: An accurate method of quantifying RSA. *Psychophysiology,*

23, 414. (abstract)

Porges, S.W. (1986b). Respiratory sinus arrhythmia: Physiological basis, quantitative methods, and clinical implications. In P. Grossman, K.H. Janssen, & D. Vaitl (Eds.), *Cardiorespiratory and cardiosomatic psychophysiology* . New York: Plenum.

Porges, S.W., McCabe, P.M., & Yongue, B.G. (1982). Respiratory-heart rate interactions: Psychophysiological implications for pathophysiology and behavior. In J. Caccioppo & R. Petty (Eds.), *Perspectives in cardiovascular psychophysiology* . New York: Guilford.

Richards, J.E. (1986). Power spectral analysis quantification of respiratory sinus arrhythmia. *Psychophysiology, 23,* 414. (abstract)

Richards, J.E. (1988). Heart rate changes and heart rate rhythms, and infant visual sustained attention. In. P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology* (Vol 3). Greenwich, CT: JAI Press.

Richards, J.E., & Casey, B.J. (1992). The development of sustained visual attention in human infants. In B. Campbell, H. Hayne, & R. Richardson (Eds.), *Information processing in infants and adults: Perspectives from human and animal research.* Hillsdale, NJ: Erlbaum Press.

Rother, M., Zwiener, U., Eiselt, M., Witte, H., et al., (1987). Differentiation of healthy newborns and newborns-at-risk by spectral analysis of heart rate fluctuations and respiratory movements. *Early Human Development, 15,* 349-363.

Schechtman, V.L., Kluge, K.A., & Harper, R.M. (1988). Time-domain system for assessing variation in heart rate. *Medical and Biological Engineering and Computing, 26,* 367-373.

Wastell, D.G. (1981). Measuring heart rate variability: Some comments on the successive difference mean square statistic. *Psychophysiology, 18,* 88-90.

Watanabe, K., Iwase, K., & Hara, K. (1973). Heart rate variability during sleep and wakefulness in low-birthweight infants. *Biology of the Neonate, 22,* 87-98.

Womack, B.F. (1971). The analysis of respiratory sinus arrhythmia using spectral analysis and digital filtering. *IEEE Transactions on Biomedical Engineering, 18,* 399-409.

Yeh, S, Forsythe, A., & Hon, E.H. (1973). Quantification of fetal heart beat-to-beat interval differences. *Obstetrics and Gynecology, 41,* 355-363.

Table 1. Cronbach's coefficient α for R-R interval variability.

	Number of samples: Sampling Duration (Total time)						
	5:60s (300s)	10:25s (250s)	15:15s (225s)	5:25s (125s)	15:5s (75s)	5:15s (75s)	5:5s (25s)
	HIGH FREQUENCY (0.333 to 1.250 Hz)						
SPA	.822						
FILTER	.850	.900	.903	.788	.875	.794	.739
PTF	.865	---[1]	.896	.705	.845[2]	.766	.616
	RESPIRATION FREQUENCY						
SPA	.799						
FILTER	.843	.898	.904	.759	.885	.816	.737
PTF	.867	---[1]	.875[2]	.811	---[1]	.758	.693
	BROAD-BAND FREQUENCY						
HPSD	.808	.872	.876	.773	.821	.643	.619
MSSD (> 0.55 Hz)	.863	.894	.885	.691	.842	.810	.763
BPF (.030 to 1.30 Hz	.834	.885	.892	.749	.858	.787	.725

[1]Number of subjects with appropriate data was less than 50%

[2]Number of subjects with appropriate data was less than total, but greater than 50%

Table 2. Significance of age effect for R-R interval measures.

	Number of samples: Sampling Duration (Total time)						
	5:60s (300s)	10:25s (250s)	15:15s (225s)	5:25s (125s)	15:5s (75s)	5:15s (75s)	5:5s (25s)
	HIGH FREQUENCY (0.333 to 1.250 Hz)						
SPA	.062						
FILTER	.030	.026	.053	.014	.036	.171	.417
PTF	.032	---[1]	.312	.902	---[1]	---[1]	---[1]
	RESPIRATION FREQUENCY						
SPA	.027						
FILTER	.045	.037	.060	.020	.035	.183	.334
PTF	.042	---[1]	.109[2]	.333	---[1]	---[1]	---[1]
	BROAD-BAND FREQUENCY						
HPSD	.043	.012	.034	.009	.049	.311	.610
MSSD (> 0.55 Hz)	.107	.087	.121	.057	.044	.215	.334
BPF (.030 to 1.30 Hz	.041	.034	.081	.026	.037	.272	.426

[1]Number of subjects with appropriate data was less than 50%

[2]Number of subjects with appropriate data was less than total, but greater than 50%

Figures

Figure 1. The reliability of R-R interval variability as a function of the sampling duration for high and respiration frequencies, and broadband frequency R-R interval variability, averaged across all measures for each frequency type.

Figure 2. The probability of the age effect on R-R interval variability as a function of the mean reliability level for the high, respiration and broadband frequencies, averaged across all measures for each frequency type.

Notes

*This research was supported by a grant from the National Institute of Child Health and Human Development, HD18942, a Small Instrumentation Grant from the National Institutes of Health, and a Research Scientist Development Award from the National Institute of Mental Health, K02-MH00958. Requests for reprints should be sent to John E. Richards, Department of Psychology, University of South Carolina, Columbia, SC 29208.

Note 1. Several other measures of R-R interval variability were also quantified and analyzed, but are not reported. Many of these variables had similar relations to the variables included in the paper. For example, both the observed respiration frequency and a fixed band were used to calculate the peak-to-trough filter (PTF) and spectral analysis (SPA) variables. The measures derived from the fixed band and from the observed respiration frequency were highly correlated ($r > .95$). A band-pass filter approach, using trigonometric-derived coefficients for the smoothing algorithm (Richards, 1986; Richards & Casey, 1991) had nearly identical results for reliability to the band-pass filter (BPF) measure developed by Porges (i.e., "V"; Porges, 1985, 1986a, 1986b).

Also, low (0.038 to 0.97 Hz) and mid (0.117 to 0.253 Hz) frequency R-R interval variability values were calculated. The short duration sampling (< 25 s) for low, and mid-frequency variability had very poor reliability. In those cases reliability was below .7. Few low frequency cycles of R-R intervals can occur in 1 minute relative to those that can occur in high frequency cycles. The smaller sampling durations (5:15s and 5:5s) affect the reliability of low and mid-frequency R-R interval variability measures more, most likely because of the inability to detect the cycling of low frequency variation over such short ranges.

Replies to reviewers:

All of reviewer B's revision suggestions were done.

Reviewer A:

(Points 1, 5). A new paragraph was done in the discussion regarding the sampling methods (multiple samples) and durations. This pointed out explicitly that both sampling duration and sampling method affect reliability and significance levels. Point 1--the drop at 125--I don't really have any speculation. I suspect it is just the particular samples that were chosen. It doesn't follow the pattern of all of the other examples given in the discussion paragraph, and in Table 1 and 2. Point 5-comparing an uninterrupted 75s or 125 s recording with 5:15s or 5:25 s. One might infer from the results and Tables and discussion that multiple samples are better than single samples. However, it is not possible to calculate a reliability coefficient on a single uninterrupted sample, so a discussion of this would seem unwarranted in this paper.

Point 2: The reviewer is technically correct, but misses the point: 1) most MSSD methods in R-R variability do event-to-event differences (e.g., beat-to-beat) rather than time-sampled MSSD, so the reviewer's points are irrelevant; 2) even with time-sampled MSSD, the cutoff frequencies are equivalent across all subjects and conditions, but Wastell points out that the changing nature of HRV power concentration across subjects and conditions makes the proportion of variance passed by the MSSD filter different across subjects and conditions; 3) I presented the approximate cut-off frequencies for the MSSD filter given the average HR of the infants in the study and the beat-to-beat MSSD method.

Point 3: Sentence added to results.

Point 4: Recommendations are made in the second and third paragraphs of the discussion regarding the amount of data required during various sampling conditions.