Psychophysiology, 40 (2003), 358–369. Blackwell Publishing Inc. Printed in the USA.

Neuroendocrine and cardiovascular reactivity to stress in mid-aged and older women: Long-term temporal consistency of individual differences

MARY H. BURLESON,[a] KIRSTEN M. POEHLMANN,[b] LOUISE C. HAWKLEY,[c] JOHN M. ERNST,[d] GARY G. BERNTSON,[e] WILLIAM B. MALARKEY,[f] JANICE K. KIECOLT-GLASER,[g] RONALD GLASER,[h] AND JOHN T. CACIOPPO[i]

[a]Department of Social and Behavioral Sciences, Arizona State University West, Phoenix, Arizona, USA
[b]Department of Psychology, University of Houston, Houston, Texas, USA
[c]Department of Psychology, University of Chicago, Chicago, Illinois, USA
[d]Department of Psychology, Illinois Wesleyan University, Bloomington Illinois, USA
[e]Department of Psychology, The Ohio State University, Columbus, Ohio, USA
[f]Department of Medicine, Division of Endocrinology, Institute for Behavioral Medicine Research, Comprehensive Cancer Center, The Ohio State University College of Medicine, Columbus, Ohio, USA
[g]Department of Psychiatry, Institute for Behavioral Medicine Research, Comprehensive Cancer Center, The Ohio State University College of Medicine, Columbus, Ohio, USA
[h]Department of Molecular Virology, Immunology, and Medical Genetics, Institute for Behavioral Medicine Research, Comprehensive Cancer Center, The Ohio State University College of Medicine, Columbus, Ohio, USA
[i]Department of Psychology, Institute for Mind and Biology, University of Chicago, Chicago, Illinois, USA

Abstract

We report long-term temporal consistency of stress-related neuroendocrine and cardiovascular variables in mid-aged and older women who performed mental math and speech stress tasks two times approximately 1 year apart. Epinephrine, norepinephrine, ACTH, cortisol, cardiac preejection period (PEP), respiratory sinus arrhythmia, heart rate (HR), blood pressure, and respiration rate were measured at baseline, after or during stressors, and 30 min posttask. Although there were exceptions, year-to-year Spearman coefficients showed mostly moderate to high consistency ($rs \approx .5$–.8) for baseline, stressor, and posttask values. For reactivity, HR and PEP were most consistent ($rs \approx .65$); consistency for other variables was moderate to low ($rs \approx .1$–.4). Means of most variables changed from year to year. Results support the use of baseline, stressor, and posttask values in longitudinal studies.

Descriptors: Consistency, Reliability, Reproducibility, Cardiovascular reactivity, Neuroendocrine reactivity, Stress

The reliability of physiological measurements across time is important for both conceptual and measurement-related reasons. Particularly in the context of behavioral medicine, a major assumption underlying research on physiological responses to psychological stress is that these responses are stable characteristics of the individuals being measured. For example, the reactivity hypothesis of cardiovascular risk states generally that individuals with exaggerated cardiovascular responses to behavioral and psychosocial stressors are at higher risk for developing cardiovascular morbidity and mortality (e.g., Krantz & Manuck, 1984). For this to be the case, the tendency to hyperrespond must be stable and reproducible over time (Manuck, Kasprowicz, & Muldoon, 1990), because it takes years for pathology to develop (Gerin, Rosofsky, Pieper, & Pickering, 1994).

The reactivity hypothesis has generated much research documenting the long-term consistency of individual differences (i.e., traits) in both tonic and phasic cardiovascular responses to stress. However, the consistency of neuroendocrine responses to stress has not been investigated to the same extent, especially in older adults prior to the onset of chronic disease. This is an

We thank Paul Wilkins, Jason Davis, Dan Litvack, David Lozano, Susan Moseley, Carolyn Cheney, Julianne Dorne, Catherine Bremer, and Tricia Rigel for their excellent technical support. We also thank the personnel of the General Clinical Research Center, including Tomasina Wall, Dana Ciccone, Bob Rice, Dave Phillips, and the nursing staff headed by Teresa Sampsel for their excellent assistance and cooperation.

This work was supported in part by the training grant MH-18831 and grants MH-42096 and MH-50538 from the National Institutes of Mental Health; program project grant AG-11585 from the National Institute on Aging; a National Institute of Health grant to the General Clinical Research Center, M01 RR00034; and The Ohio State University Comprehensive Cancer Center core grant, CA 16058.

Address reprint requests to: Mary H. Burleson, Department of Social and Behavioral Sciences, Arizona State University West, 4701 West Thunderbird, Phoenix, Arizona 85069, USA. E-mail: mary.burleson@asu.edu.

especially important gap in the literature because individual differences in stress reactivity of both cardiovascular and neuroendocrine responses are thought to contribute to the development of chronic disease in this population. Psychological stress affects the sympathetic-adrenal-medullary (SAM) axis and the hypothalamic-pituitary-adrenal (HPA) axis, as well as the parasympathetic branch of the autonomic nervous system (ANS). In the current study, we use Spearman rank correlations to estimate the year-to-year consistency of individual differences in baseline, stressor, posttask, and change scores for a number of variables often employed by researchers studying the effects of stress on these systems.

Many neuroendocrine studies of the stress response have focused on the catecholamines norepinephrine (NEPI) and epinephrine (EPI), as indicators of activity in the SAM axis, or on cortisol, and sometimes adrenocorticotropic hormone (ACTH), to indicate HPA activity. In previous reports of the temporal consistency of plasma or serum catecholamines, retest intervals ranged from 1 hr to 12 weeks (Cohen et al., 2000; Grassi et al., 1997; Jern, Pilhall, Jern, & Carlsson, 1991; Mills, Berry, Dimsdale, Neleson, & Ziegler, 1993; Mills, Ziegler, Dimsdale, & Parry, 1995; Seraganian et al., 1985; Siever et al., 1986). In these studies, consistency estimates for baseline and stressor values of catecholamines tended to be higher than for change scores (e.g., Pearson *r*s of .23–.88 vs. .02–.58, respectively), likely because measurement error is compounded in the calculation of delta scores (Strube, 1990). Not surprisingly, for both catecholamines, the highest consistencies (*r*s = .87 for EPI; .88 for NEPI) were seen at the shortest retest interval (1 hr; Jern et al., 1991). The temporal consistency of blood levels of EPI and NEPI are not identical, but no clear pattern can be seen in which one is clearly more reliable than the other.

Prior studies of the reliability of cortisol have been performed using saliva (e.g., Cohen et al., 2000; Coste, Strauch, LeTrait, & Bertagna, 1994; Houtman & Bakker, 1991; Kirschbaum et al., 1995, 1990; Kirschbaum, Wust, Faig, & Hellhammer, 1992; Pruessner et al., 1997; Smyth et al., 1997) or plasma or serum (Coste et al., 1994; Huizenga et al., 1998; Schulz & Knabe, 1994; Thomas et al., 1994; Zorilla, DeRubeis, & Redei, 1995). For the most part, these studies have assessed basal levels rather than stress responses, with retest intervals ranging from one day to three months. One retest interval was 2.5 years (Huizenga et al., 1998). In general, reliability for baseline levels has ranged from low to moderately high for cortisol in both blood and saliva (*r*s = .16–.70); the highest values (*r* = .85–.89) were obtained in twins (Kirschbaum et al., 1992; Young, Aggen, Prescott, & Kendler, 2000). Reliability of stress-task and reactivity values for cortisol in blood has not been reported. Reliability of stress-task values of salivary cortisol was moderately low in the single study reporting it (*r*s = .35, .48; Houtman & Bakker, 1991). Reliability of reactivity has been assessed using baseline-adjusted residual change scores (*r* = .37; Cohen et al., 2000) and area under the curve (Spearman coefficients .17–.60; Kirschbaum et al., 1995). Only one study was found that measured the temporal reliability of ACTH (Coste et al., 1994). Moderately low baseline reliability was reported [intraclass correlation coefficient (ICC) = .48] in 20 young male participants over a 5-week interval.

As noted above, numerous studies and reviews have documented the consistency of stress-related heart rate and blood pressure (BP; e.g., Manuck, Kamarck, Kasprowicz, & Waldstein, 1993; Swain & Suls, 1996). These variables often are used as indicators of underlying changes in ANS activity, to which alterations in both the sympathetic and parasympathetic branches can contribute independently (Berntson, Cacioppo, & Quigley, 1991). Efforts to pinpoint the autonomic origins of changes in HR require separate measures of the two branches. Sympathetic activation of the heart has been assessed using cardiac preejection period (PEP; Berntson et al., 1994; Cacioppo et al., 1994), and parasympathetic activation has been assessed using respiratory sinus arrhythmia (RSA; Berntson et al., 1994; Cacioppo et al., 1994), high frequency heart rate variability (HRV), and heart period variability (HPV; see Berntson et al., 1997).

Although its psychometric properties are not as well known as those of HR and BP, numerous reports for both tonic levels and reactivity of PEP have been published (Allen, Sherwood, Obrist, Cromwell, & Grange, 1987; Burns, Ferguson, Fernquist, & Katkin, 1992; Kamarck et al., 1992; Kamarck, Jennings, Stewart, & Eddy, 1993; Kasprowicz, Manuck, Malkoff, & Krantz, 1990; Llabre et al., 1993; Sherwood et al., 1997; Sherwood, Turner, Light, & Blumenthal, 1990; Willemsen et al., 1998). Retest intervals in these studies have ranged from 4 weeks to 10 years. Generally, consistency values for baseline and stress-task levels were moderate to high (*r*s = .60–.93). The one value that fell below this range employed one of the longest retest intervals (2.5 years), and used different methods to estimate the PEP at test and retest (Allen et al., 1987). Consistency for reactivity values was more variable in the studies above (*r*s = .15–.86).

Most studies of the reliability of RSA or other measures of parasympathetic activation in healthy adults have assessed these variables during basal conditions (e.g., Byrne, Slater, & Porges, 1991; Hatch, Borcherding, & Norris, 1990; van de Borne, Montano, Zimmerman, Pagani, & Somers, 1997) or over a 24-hr period (e.g., Klieger et al., 1991; Kochiadakis et al., 1997; Pitzalis et al., 1996). Retest intervals have ranged from 10 min to approximately 8 months. Values derived from these studies are moderately to highly consistent (ICCs = .48–.95; *r*s = .55–.86). The single study of laboratory stress reactivity found low reliability (ICC = .35 for aggregated arithmetic and reaction time tasks) for reactivity of high frequency heart period variability in mid-aged adults (Sloan, Shapiro, Bagiella, Gorman, & Bigger, 1995).

To summarize, past estimates of temporal consistency for physiological indicators of stress vary widely, suggesting that the psychometric properties of stress reactivity measures depend on the nature of the acute stressor and the analytical method for measuring reactivity as well as the specific physiological response system in question. The cardiovascular variables PEP and RSA have shown moderate to high baseline reliability over intervals of up to several years. In contrast, the reliability of circulating baseline plasma catecholamine values has varied from moderate to poor with the longest retest interval of only 1 week. For both neuroendocrine and cardiovascular measures, baseline and stress-task levels appear in general to be more reliable than do measures of reactivity in these variables.

In this study, we provide the first report of 1-year temporal consistency of individual differences in stress reactivity of cortisol and ACTH, using a stress reactivity protocol that ensures the task is comparably demanding for all individuals (cf. Cacioppo et al., 1995). In addition, we extend previous work on the temporal consistency of tonic and phasic levels of neuroendocrine and cardiovascular variables by studying an older female population that has not previously been tested and by using a long interval between tests. We measured the consistency over

time of neuroendocrine and cardiovascular measures in a sample of healthy mid-aged and older women, who performed laboratory speech and math stress tasks on two occasions approximately 1 year apart. Measures were taken before, during or immediately following, and 30 min after stressor exposure during both assessments. Physiological parameters assessed included plasma levels of EPI, NEPI, ACTH, and cortisol; PEP, RSA, HR, SBP, and DBP. Respiration rate was included because of its importance as a control measure when assessing cardiovascular reactivity.

Method

Participants

Participants in the current study were obtained from a larger sample of women assessed in a longitudinal study of psychological and physiological effects of chronic stress that began 7 years prior to our data collection. They were initially recruited by advertisement in the community. In the first year of the current study, there were 45 participants; a subset of 35 returned for a second assessment the following year. The participants ranged in age from 49 to 83 (M = 67.8, SE = 1.2) in Year 1. Thirty-eight were Caucasian and 7 were African American. Mean body mass index (BMI; calculated as weight in kilograms divided by squared height in meters) at entry was 25.9 (SE = 0.7). All were postmenopausal, and 36% received estrogen replacement therapy throughout the study. All participants met the following inclusion criteria at both measurement occasions: (a) no chronic disease; (b) no current illness; (c) on average, less than 10 hr of exercise per week; (d) on average, less than 10 alcoholic beverages per week; (e) no math, speech, or needle phobia; and (f) consistent usage or nonusage of beta-adrenergic receptor blockers throughout the study.

The Year 1 characteristics of the retest sample of 35 women were compared to those of the women who were not retested in Year 2. The two groups did not differ in race, hormone use, marital status, BMI, activity level, alcohol or caffeine consumption, sleep, or fear of speech or math. However, the returnees were older than the dropouts, $F(1,44) = 5.14$, $p = .028$, M (SE) = 69.1 (1.3) and 63.0 (2.4) years, respectively. The returnees also reported significantly less fear of needles and greater effort to complete the math stress task. Comparison of the two groups' cardiovascular and neuroendocrine responses revealed no differences. Thus, participants who were younger, more fearful of needles, and thought the math stress task required less effort were less likely to return for reassessment.

Procedures

To prepare for their appointments, participants were asked to reschedule if they experienced illness or a major negative life event. They also were asked to refrain from (a) using alcohol or nonprescription medication or exercising the day before the study and (b) eating, drinking anything besides water, or smoking from midnight until the time of their scheduled appointment the following morning.

When a participant arrived (at 8:00 a.m.), the stress tasks and measures were reviewed, any questions were answered, and informed consent was obtained. A strain-gauge respirometer was placed around the lower chest, and disposable band electrodes for impedance cardiography were placed around the participant's neck and chest. The participant was then asked to lie down and a 20-gauge catheter was inserted into an antecubital vein. After catheter insertion, the participant's arm and hand were placed on a heating pad to arterialize the venous blood flow, and a wrist tonometer was placed over the radial artery of the other arm for beat-to-beat blood pressure measurement. To allow adaptation to the setting, the participant rested in a seated position while a set of questionnaires was verbally administered for approximately 30 min. The scales that were used and the results of those assessments will be reported elsewhere. After the adaptation period, participants were asked to rest quietly and relax while cardiovascular and respiratory measures were recorded for 6 min.To assess baseline neuroendocrine function, the first blood sample was collected immediately following the cardiovascular recording. Because physiological parameters are often influenced by a diurnal rhythm, we monitored the year-to-year difference in timing between corresponding pairs of blood samples (reported below).

Following the baseline measures, participants received instructions for the two stress tasks (described below), and any questions about the stressors were answered. The participants then performed the stress tasks. The second stressor immediately followed completion of the first, and the order of these stressors was counterbalanced across participants and years to allow examination of order effects and to reduce habituation. Cardiovascular and respiratory measures were collected throughout the 6-min speech and math stressor periods, and the second and third blood samples were drawn immediately after each stressor for neuroendocrine assays. In addition, participants were asked to rate the stressors and their responses to the stressors immediately after completion of the second stressor during the third blood draw.

After the stress tasks and third blood sample, participants were instructed to sit quietly and relax for an additional 30 min. Cardiovascular and respiratory recording occurred during the last 3 min of this period, after which the final blood sample was drawn.

Stress Tasks

Year 1. For the math stressor, participants performed 1-min serial subtraction problems aloud, continuously for 6 min. They were instructed that the experimenter would correct any errors they made, and that they should continue from the correct number. To maintain maximal stress task involvement and moderate stress task difficulty (i.e., approximately 10 correct answers per minute; Cacioppo et al., 1995; Uchino, Kiecolt-Glaser, & Cacioppo, 1992), better performance led to more difficult math problems in the subsequent minute.

For the speech stress task, we used an adaptation of the speech stressor of Saab, Matthews, Stoney, and McDonald (1989). Each participant was asked to imagine that she had been taken to see a department store manager by a security guard who had falsely accused her of shoplifting a belt, and allowed to make a speech in her defense that covered a specific set of points. Participants were instructed to give intelligent and well-thought-out answers because their speeches would be recorded and compared with the speeches of others. They had 3 min to prepare and 3 min to present their speeches.

Year 2. The math stress task, the difficulty of which was contingent on the participant's performance, was the same as in Year 1. The speech task was structured in the same way but used

a different topic in order to reduce habituation and practice effects. The participant was asked to imagine that she was falsely accused of hitting another car with her car, and to prepare a talk to deliver to the police, covering a set of specific points equivalent to those used in Year 1. All other instructions were the same as in Year 1.

Neuroendocrine Measures

Cortisol and ACTH were tested in heparinized plasma stored at -70° until assay. Cortisol was measured with a fluorescent polarization technique (TDX-Abbott Lab, Chicago, IL) with intra- and interassay coefficients of variation (CVs) of 10% or less. ACTH was measured using an immunoradiometric assay (Allegro HS-ACTH kit, Nichols Institute) with a sensitivity of 1 pg/ml, an intraassay CV of 3% and an interassay CV of 8% or lower.

Levels of EPI and NEPI were determined in plasma treated with ethylenediaminetetra-acetic acid (stored at -70°C until assay), using high-performance liquid chromatography on a Waters system (Millipore, Waters Division, Marlborough, MA) with electrochemical detection. Alumina was used for extraction; extraction efficiency, evaluated with a dihydroxybenzylamine (DHBA) standard, is 60%–90%. A Waters catecholamine eluent was used for the mobile phase. For EPI, the sensitivity of this system is 10 pg/ml, and the intra- and interassay CVs are 12% or less. For NEPI, sensitivity is 20 pg/ml and CVs are 9% or less.

Cardiovascular Measures

A Minnesota Impedance Cardiograph (Model 304B; Instrumentation for Medicine, Greenwich, CT) was used to measure the electrocardiogram (ECG), basal thoracic impedance (Z_0), and the first derivative of the change in the impedance signal (dZ/dt). Disposable band electrodes (Instrumentation for Medicine) were placed around the neck and chest in a tetrapolar configuration (see Sherwood, Allen, et al., 1990). A 4-mA alternating current at 100 kHz was passed through the outer two electrodes, and Z_0 and dZ/dt were recorded from the two inner electrodes. SBP and DBP were collected using the radial tonometer on the Colin Pilot 9201 blood pressure monitor (Colin Medical Instruments, San Antonio, TX), which yields a continuous pulsatile waveform. Respiration was recorded using an EPM Systems amplifier and strain gauge (EPM Systems, Midlothian, VA) placed around the trunk below the lowest current electrode. The impedance cardiograph, BP monitor, and respirometer were interfaced with a microcomputer, and the Z_0, dZ/dt, ECG, SBP, DBP, and respiratory signals were converted to digital signals (12-bit A/D converter; 500 Hz for dZ/dt and ECG, 250 Hz for Z_0 and respiration, and 5-beat averages for BP) which were edited, reduced, and analyzed off-line. To minimize artifacts in recording interbeat intervals, the ECG and dZ/dt waveforms were monitored during collection and ECG was bandpass filtered (1 Hz to 10,000 Hz) prior to digitization.

The impedance data were ensemble averaged within 1-min epochs, and each waveform was verified or edited prior to analyses using interactive software (Kelsey & Guethlein, 1990). Minutes were rejected if artifacts comprised more than 10% of the beats. The PEP was quantified as the time interval in milliseconds from the onset of the ECG Q-wave to the B-point of the dZ/dt wave.

For RSA, the interbeat intervals were edited and artifacts removed using the algorithm of Berntson, Quigley, Jang, and Boysen (1990), then transformed to a 500-ms interval time series. Minutes were rejected if artifacts comprised more than 10% of the beats or if 30 consecutive s of clean data could not be isolated. Using a PC-based software package (MXedit 2.01, Delta-Biometrics, Bethesda, MD), a 21-point cubic digital filter was moved stepwise through the time series to remove nonstationarity (Porges & Bohrer, 1990). RSA was quantified as the natural logarithm of the variance in the residual time series occurring within the frequency bandpass associated with respiration (0.12–0.40 Hz).

BP measures were derived from the Colin tonometric data using custom software to mark peaks and valleys to indicate SBP and DBP, respectively.

Because respiration rate is often used for correction and interpretation of RSA values, we measured it during cardiovascular data collection. The respiration signal was bandpass filtered (0.12–0.40 Hz) using an interpolated finite impulse response filter, digitized, then verified or edited to eliminate movement artifacts.

Minute-by-minute means were calculated for each variable, then averaged over the 6-min baseline, each 6-min stressor, and the 3-min posttask period.

Data Analysis

Year-to-year temporal consistency was evaluated using Spearman correlations (to assess the consistency of the rank ordering from year to year) for each measurement period. To increase external validity, responses to the math and speech stressors were aggregated, and consistency estimates were derived for speech, math, and the aggregated stressor response. Reactivity values were delta scores derived by subtracting the baseline value from the mean of the aggregated math and speech stressor values. Reactivity and recovery within each year and changes in mean values from year to year within each time point were evaluated with separate repeated measures analyses of variance (ANOVAs).

We had some concern about small sample sizes and missing data because 10 participants did not return for Year 2 assessment and other missing values occurred for most of the variables. Therefore, we followed the recommendations of Graham, Hofer, Donaldson, MacKinnon, and Schafer (1997) and used full information maximum likelihood estimation to derive year-to-year consistency coefficients for all of the measures reported. For the most part, these standardized covariance estimates were very similar in magnitude and significance to the Spearman correlations we report, suggesting that missing data did not have a major effect on the estimates. Further, although sample sizes were small for some variables, power to detect correlations of .4 or above was at least 0.7 in all cases.

Preliminary analyses. RSA was evaluated using both raw values and values that were corrected for respiration rate. Parameter estimates were virtually identical for the two methods. Further, there were no year-to-year differences in respiration rate. Therefore, the values for raw, rather than corrected, RSA are reported here.

The average time for starting physiological data collection was 10:11 a.m. ($SE = 4$ min) in Year 1 and 10:10 a.m. ($SE =$ 2 min) in Year 2. Examination of blood sample timing revealed good consistency from Year 1 to Year 2. Means and standard errors (in minutes) of the absolute values of the year-to-year differences for corresponding pairs of blood samples were as follows: at baseline (17.4 [2.5]); after first stressor (18.3 [2.4]);

after second stressor (18.1 [2.5]); after posttask recording (19.5 [2.4]).

Results

Stress Task and Recovery Effects

To confirm that the stressors resulted in activation of the stress response, neuroendocrine and cardiovascular data were analyzed within each year using the baseline and aggregated stressor scores. Figure 1 illustrates the Year 1 and Year 2 mean values for baseline, aggregated stressor, and posttask time periods for the four hormones measured. Corresponding values for the cardiovascular and respiratory variables are found in Figure 2. Table 1 provides Year 1 and Year 2 means and *SE*s for baseline, stressor, and posttask values, along with effect sizes for within-year reactivity and recovery analyses, for all of the measured variables. Means and SEs for separate speech and math stressor responses are also provided.

As expected, aggregated stressor values of EPI, NEPI, HR, SBP, DBP, and respiration rate were significantly higher than baseline values in both Year 1 and Year 2. RSA aggregated stressor values were significantly lower than baselines for both years. For ACTH, reactivity was significant for Year 1 but not Year 2. Cortisol and PEP reactivity were not significant for either year.

The only variables that were fully recovered (equivalent to baseline) at the posttask measurement in both years were cortisol and DBP. Posttask EPI and ACTH did not differ from baseline in Year 1 but were higher than baseline in Year 2. Posttask NEPI was higher and posttask SBP was lower than baseline in Year 1 but neither differed from baseline in Year 2. HR, PEP, and breaths per minute were higher at posttask than baseline in both years. Finally, posttask RSA was lower than baseline in both years.

Temporal Consistency

Table 2 provides year-to-year Spearman correlations for all variables.

Baseline, stressor, and posttask values. For the neuroendocrine measures, all of the year-to-year Spearman correlations for baseline, speech, math, and aggregated stressor values were highly significant. Spearman correlations for the posttask values were smaller, yet still significant. Across all of these time points, NEPI was most consistent from year to year.

For the cardiovascular and respiratory variables, many Spearman correlations for baseline, speech, math, and aggregated stressor values were highly significant, and all were significant at $\alpha = .05$. Spearman correlations for the posttask time points were highly significant for all of these variables except DBP. Across all time points, HR was most consistent from year to year.

Reactivity scores. In the current study, Spearman coefficients for consistency of reactivity of all of the measured hormones were relatively low; the estimate for EPI was the only significant value.

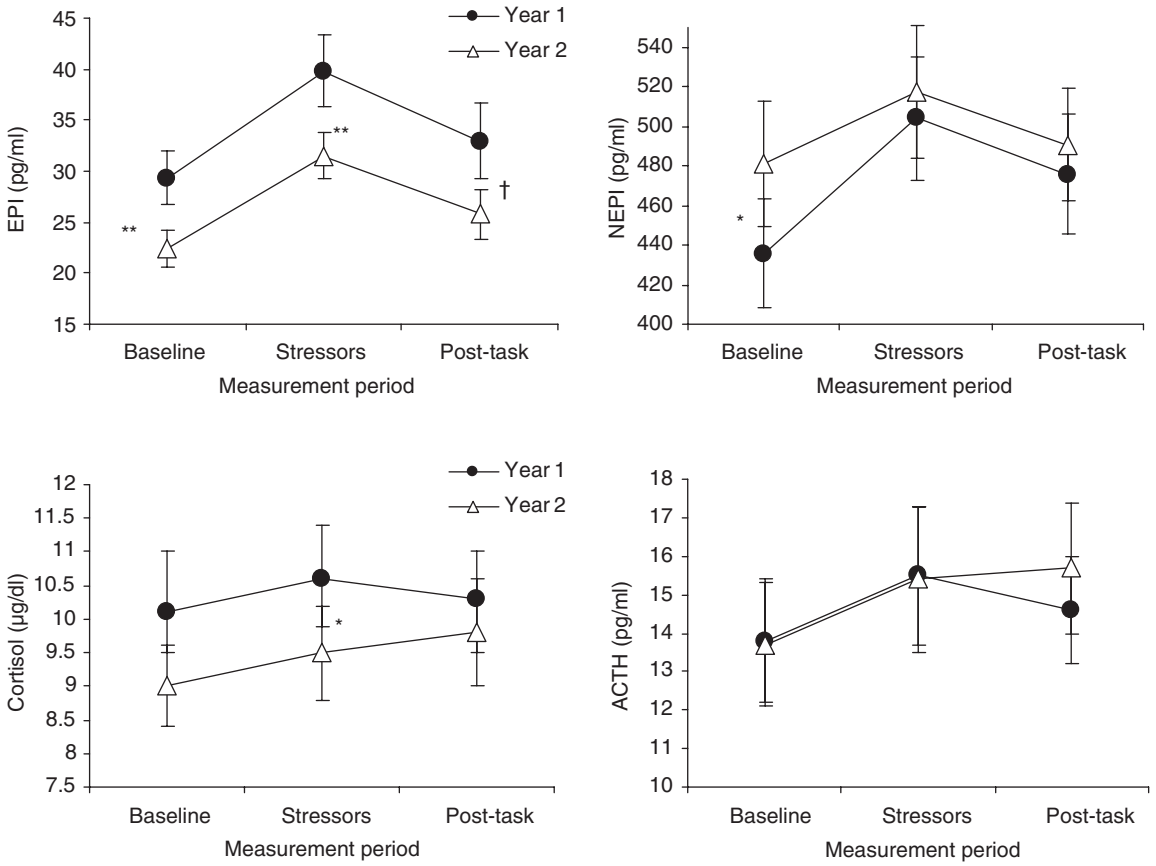

Figure 1. Mean levels of neuroendocrine variables for Year 1 and Year 2. Vertical lines depict standard errors of the means. Baseline was measured in plasma drawn immediately after 6-min baseline period. Stressors: mean of measurements from plasma drawn immediately after speech stress task and immediately after math stress task. Posttask was measured in plasma drawn 30 min after final stress task was completed. Symbols indicate significance of differences between Year 1 and Year 2 means. $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

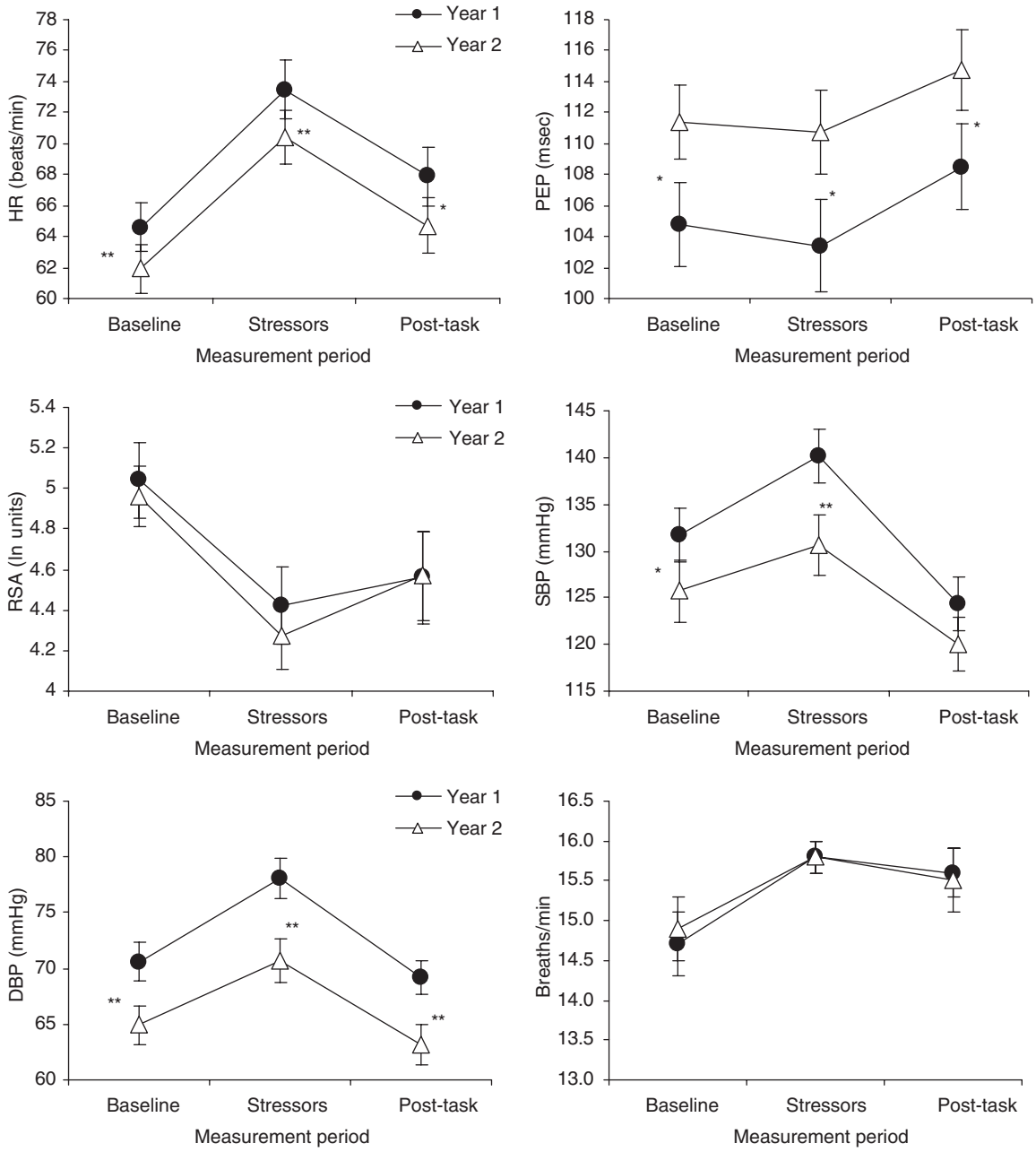

Figure 2. Mean levels of cardiovascular and respiratory variables for Year 1 and Year 2. Vertical lines depict standard errors of the means. Baseline was recorded during 6-min baseline period. Stressors: mean of measurements recorded during 6-min speech stress task and 6-min math stress task. Posttask was recorded during final 3 min of 30-min posttask waiting period. Symbols indicate significance of differences between Year 1 and Year 2 means. *$p < .05$. **$p < .01$.

The Spearman correlation for cortisol reactivity was essentially zero.

For the cardiovascular and respiratory variables, Spearman correlations for delta values were highly significant for HR and PEP, whereas deltas for the other variables were not significantly correlated from year to year. The Spearman correlation for RSA reactivity was essentially zero.

Changes in Means from Year to Year

As noted above, Figures 1 and 2 illustrate the yearly means and SEs for baseline, aggregated stressor, and the posttask time periods for all variables. Significant changes from year to year are noted in the figures.

Baseline NEPI was significantly higher in Year 2. All other significant changes in baseline, stressor, or posttask scores from Year 1 to Year 2 indicate lower levels of physiological response (longer PEPs reflect lower sympathetic activation). EPI, HR, SBP, and DBP showed consistent declines from Year 1 to Year 2, whereas PEP increased, across baseline, speech, math, and aggregated posttask values. NEPI, ACTH, cortisol, RSA, and respiration rate were somewhat more stable. Although Year 2 changes were not as robust for posttask values, they showed a

Table 1. *Means (Standard Errors) for Baseline, Speech, Math, Aggregated Stressor, and Posttask Values, and Effect Sizes for Reactivity and Recovery*

Measure	Year	*n*	Base	Spch[a]	Math[a]	Stress	$F_{reactivity}$	D^b	Post	$F_{recovery}$	D^c
					Neuroendocrine variables						
EPI (pg/ml)	Year 1	31	29.3	41.1	38.5	39.8	47.29**	.72	32.5	1.90	.22
			(2.6)	(3.8)	(3.4)	(3.5)			(3.6)		
	Year 2	34	22.1	31.7	31.4	31.6	47.66**	1.07	25.9	10.70**	.43
			(1.6)	(2.2)	(2.3)	(2.1)			(2.2)		
NEPI (pg/ml)	Year 1	29	435.8	504.6	503.3	503.9	38.32**	.46	479.4	14.90**	.30
			(27.4)	(29.9)	(33.2)	(30.9)			(29.8)		
	Year 2	34	517.8	570.7	544.8	557.8	9.20**	.20	536.1	1.15	.09
			(34.4)	(36.6)	(34.3)	(34.7)			(31.6)		
ACTH (pg/ml)	Year 1	28	13.5	15.9	14.4	15.1	5.32*	.20	15.1	1.79	.20
			(1.6)	(2.0)	(1.8)	(1.9)			(1.5)		
	Year 2	32	12.7	14.5	14.3	14.3	2.90†	.22	15.5	5.17*	.38
			(1.3)	(1.8)	(1.6)	(1.7)			(1.5)		
Cortisol (μg/dl)	Year 1	31	10.1	10.4	10.8	10.6	2.21	.11	10.4	0.22	.06
			(0.9)	(0.8)	(0.9)	(0.8)			(0.7)		
	Year 2	34	9.1	9.0	9.6	9.3	0.29	.05	9.7	0.51	.17
			(0.6)	(0.6)	(0.7)	(0.6)			(0.7)		
					Cardiovascular and respiratory variables						
HR (bpm)	Year 1	32	64.7	74.6	72.4	73.5	94.38**	.94	67.9	23.81**	.34
			(1.6)	(2.0)	(2.0)	(2.0)			(1.9)		
	Year 2	35	61.9	71.3	69.4	70.4	160.73**	.91	64.7	21.98**	.30
			(1.6)	(1.8)	(1.7)	(1.7)			(1.7)		
PEP (ms)	Year 1	29	104.6	100.8	102.8	101.8	3.23†	−.20	108.1	5.80*	.24
			(2.7)	(3.2)	(3.2)	(3.2)			(2.9)		
	Year 2	31	110.8	109.6	111.0	110.3	0.24	−.04	114.7	13.15**	.32
			(2.2)	(2.6)	(2.5)	(2.5)			(2.4)		
RSA (ln units)	Year 1	31	4.99	4.27	4.48	4.37	20.37**	−.60	4.56	12.26**	−.42
			(0.19)	(0.18)	(0.21)	(0.19)			(0.22)		
	Year 2	32	4.92	4.20	4.32	4.26	54.76**	−.75	4.59	8.38**	−.38
			(0.16)	(0.16)	(0.19)	(0.17)			(.21)		
SBP (mmHg)	Year 1	29	130.3	145.5	138.9	142.2	26.83**	.77	124.7	13.26**	−.36
			(2.9)	(3.4)	(2.9)	(3.0)			(3.3)		
	Year 2	34	124.2	132.0	130.2	131.1	9.73**	.40	121.5	0.85	−.15
			(3.2)	(3.6)	(2.7)	(2.9)			(2.8)		
DBP (mmHg)	Year 1	32	69.9	79.0	77.3	78.2	32.47**	.90	68.57	0.83	−.15
			(1.7)	(1.9)	(1.9)	(1.7)			(1.5)		
	Year 2	34	64.5	70.6	72.4	71.5	26.66**	.67	64.1	0.08	−.04
			(1.8)	(2.2)	(2.1)	(1.9)			(1.8)		
Breaths per min	Year 1	33	14.8	15.7	16.0	15.9	8.46**	.51	15.6	5.15*	.36
			(0.4)	(0.3)	(0.2)	(0.2)			(0.3)		
	Year 2	34	14.9	15.7	15.9	15.8	6.88*	.44	15.5	7.63**	.28
			(0.4)	(0.3)	(0.3)	(0.2)			(0.4)		

Notes: Base: baseline; spch: speech task; math: math task; stressor: aggregated speech and math tasks; post: posttask; EPI: epinephrine; NEPI: norepinephrine; ACTH: adrenocorticotropic hormone; HR: heart rate; PEP: preejection period; RSA: respiratory sinus arrhythmia; SBP: systolic blood pressure; DBP: diastolic blood pressure.

[a]Order of speech and math tasks was counterbalanced between years. [b]Effect size for reactivity analysis = $M_{aggregated\ stressor} - M_{baseline}/SD_{baseline}$. [c]Effect size for recovery analysis = $M_{posttask} - M_{baseline}/SD_{baseline}$.

†$p < .10$. *$p < .05$. **$p < .01$.

similar pattern, with the exception of SBP, for which posttask values did not change significantly from Year 1 to Year 2.

Discussion

The current study reports year-to-year consistency of individual differences for baseline, stressor-related, posttask, reactivity, and recovery levels of physiological variables often studied in the context of psychological stress. Because one major goal was to assess temporal consistency of reactivity, we were very concerned about aspects of our protocol that might influence these estimates. Repeated measurements of the same responses are particularly susceptible to habituation and loss of interest, and we had two stressors that were repeated over time. As described above, both stressors were altered somewhat between measurement occasions to reduce habituation and maintain participant engagement; thus the coefficients we report cannot strictly be interpreted as test–retest reliability estimates. On the other hand, because real-life stressors are unlikely to replicate one another, this procedure may more closely resemble stress experienced in everyday life. Thus, the consistency of individual differences found here may better reflect that found in the field than would an estimate of classical test–retest reliability.

A potential criticism of the current results is that due to systematic attrition, our final sample was not representative of the initial group (they were older and less fearful of needles). If our initial group had been a random stratified sample, this would

Table 2. *Spearman Correlations for Year 1 with Year 2*

Measure	*n*	Baseline	*n*	Speech[a]	*n*	Math[a]	*n*	Stressor	*n*	Delta	*n*	Posttask
Neuroendocrine variables												
EPI (pg/ml)	31	.56**	31	.63**	31	.55**	31	.60**	31	.37*	30	.49**
NEPI (pg/ml)	29	.70**	29	.80**	29	.70**	29	.78**	29	.33†	28	.73**
ACTH (pg/ml)	29	.68**	28	.70**	29	.68**	29	.72**	29	.20	27	.43*
Cortisol (μg/dl)	31	.51**	31	.72**	31	.63**	31	.72**	31	.06	30	.41*
Cardiovascular and respiratory variables												
HR (bpm)	34	.85**	34	.76**	34	.81**	34	.79**	34	.67**	32	.78**
PEP (ms)	28	.59**	27	.45*	28	.57**	28	.54**	28	.65**	29	.54**
RSA (1n units)	32	.63**	32	.56**	32	.84**	32	.81**	32	.07	30	.59**
SBP (mmHg)	35	.64**	33	.66**	32	.41*	31	.66**	31	.32	31	.65**
DBP (mmHg)	35	.60**	33	.62**	32	.41*	31	.57**	31	.21	31	.35†
Breaths/min	34	.57**	35	.68**	35	.37*	35	.53**	34	.31†	33	.54**

Notes: Stressor: aggregated speech and math stress tasks; delta: stressor − baseline. For all nonsignificant correlations, power (1 − β) to detect correlations of at least .40 was .70 or higher.
[a]Order of speech and math stress tasks was counterbalanced between years.
$^{\dagger}p < .10$. *$p < .05$. **$p < .01$.

be of greater concern. However, even our initial group likely differed systematically from the general population, as they were a self-selected group of volunteers willing to provide personal information and repeated blood samples for a small reward. Furthermore, the dropouts did not differ from the returnees in any cardiovascular or neuroendocrine measurement. Despite potentially reduced generalizability, we believe our results are valuable because they provide the first estimates of temporal consistency for some of these variables in any group of mid-aged and older women.

Baseline and Stressor Values

In general, hormones are released in a pulsatile fashion; hence their plasma levels may change substantially within a short time frame. Our year-to-year Spearman correlations for baseline and stressor values of NEPI and EPI are remarkably high, given the susceptibility of these parameters to environmental perturbations, endogenous rhythms in their release, and the difficulty of the assays involved (Ziegler, 1989). The only comparable values from previous research were measured over much shorter time frames (e.g., 1 hr to 1 day; Grassi et al., 1997; Jern et al., 1991; Siever et al., 1986). In the current study, we used heating pads to arterialize the blood flow for sampling, likely causing the venous catecholamine levels to more closely approximate arterial levels (Veith, Best, & Halter, 1984). Compared to venous blood samples, arterial samples may provide catecholamine levels that more sensitively reflect sympathetic nervous system stress responses (Dimsdale & Ziegler, 1991; Goldstein, Eisenhofer, Sax, Keiser, & Kopin, 1987; Hjemdahl, Freyschuss, Juhlin-Dannfelt, & Linde, 1984; but see Stoney & Hughes, 2001). Thus, arterialization of our samples for NEPI and EPI determinations may have improved their year-to-year consistency to the extent that the underlying responses were themselves consistent.

Our year-to-year consistency estimate for baseline ACTH is higher than the only previous report (Coste et al., 1994), but the moderately high estimates we report for baseline cortisol are generally in agreement with previous results. Although cortisol is highly responsive to novelty (Rose, 1984), moderately consistent baseline values have been obtained over intervals as long as 2.5 years (Huizenga et al., 1998). The only previous consistency estimates for stressor levels of cortisol (Houtman & Bakker, 1991) were moderately low, in contrast to our findings. Our results support the idea that, in a controlled environment with minimal differences between measurement contexts, baseline and stressor levels of HPA function can be measured with moderate to high year-to-year consistency.

Our year-to-year Spearman correlations for baseline and aggregated stress task HR, SBP, DBP, PEP, and RSA were highly significant. Many of these results are very similar to previous consistency estimates based on comparable retest intervals (e.g., Llabre et al., 1993; Sloan et al., 1995; Stoney, Niaura, & Bausserman, 1997). However, no previous study has reported consistency of PEP or RSA in a healthy sample as old as the one used here. Given known age-related changes in heart functioning, our results for PEP are particularly encouraging. Aging may be associated with a leaky aortic valve, or with increasing levels of body fat. Both of these factors may cause changes in beat-to-beat impedance waveforms that increase measurement error for PEP (e.g., smearing of the B-point of the dZ/dt wave). Differences in electrode placement between measurement occasions can also increase measurement error. Nevertheless, our findings show that significant long-term consistency is still achievable, supporting the study of PEP and RSA in older samples.

As noted above, most of the baseline, stress task, and posttask means changed from Year 1 to Year 2 in the direction consistent with lower levels of physiological activation. It is likely that the participants' increased familiarity with the measurement setting and personnel contributed to this phenomenon.

Posttask Measures

In the context of laboratory stress research, recovery usually refers to a process by which the individual returns to a baseline state after experiencing a stressor. It is often measured as either the time required to return to the prestressor baseline state or the degree of elevation above this baseline that still persists at the end of a predetermined posttask interval (Stewart & France, 2001). Recovery is an important issue, particularly in the context of aging research (e.g., Seeman & Robbins, 1994). For example, loss of resilience in stress response systems may be a critical feature of the build-up of stress over time (McEwen, 1998). However, reliable measurement of recovery across multiple

physiological systems can be challenging because these systems likely recover at different rates, but a given protocol may demand that they are measured at the same time intervals. Thus, the selection of measurement interval can strongly influence which variables appear to demonstrate reliable recovery (Linden, Earle, Gerin, & Christenfeld, 1997). In the current study, we measured all of the neuroendocrine and cardiovascular variables approximately 30 min after completion of the stress tasks. During these 30 min, the participants sat quietly without moving or speaking, making it somewhat surprising that many of the variables were not at their baseline levels at the posttask assessment. Nevertheless, despite the fact that most of the variables had not returned to baseline, all consistency estimates for these posttask measures were significant except for DBP. These findings suggest that the processes that occur after a laboratory stress task are reproducible from year to year. Future research should incorporate more frequent samples to ensure adequate resolution in the study of the recovery process.

Measures of Reactivity

Perhaps more than any other type of measurement in stress research, the reliability of stress reactivity depends inherently on the protocol used to elicit it, making careful protocol design and execution crucial in studies of reactivity. This is the case for several reasons. First, as found in the current study, the range of individual differences in reactivity tends to be smaller than the range of differences in baseline scores, and this smaller range will tend to reduce reliability. Therefore, it is important to attend to the magnitude of the individual differences in reactivity obtained in a given protocol, not simply to the magnitude of the mean change scores across individuals. Protocols that employ interpersonal or social stressors, such as our speech task, are particularly effective for eliciting a broad range of responses (Linden, Rutledge, & Con, 1998).

The second reason involves the issue of aggregation to reduce measurement error. Although always important, aggregation is even more vital when measuring reactivity, because measurement error is usually higher when composite scores, such as deltas, are used (Strube, 1990). Aggregation can be carried out by combining multiple measurements made in the same context, as in our averaging of minutes within each measurement epoch. This type of aggregation improves consistency if the stressor has similar effects throughout the epoch on the measured response. Our protocol encouraged sustained, consistent reactivity during the math task by making the difficulty of the problems contingent on the participants' performance.

Protocols differ in the extent to which they allow combination of measures and also in the range of measurement contexts. For example, our procedure allowed us to combine the responses to two stress tasks at each occasion, and these stressors differed somewhat in the demands they placed on the participants. Aggregation of reactivity to different stressors on the same measurement occasion improves construct validity by providing a better estimate of dispositional stress responsivity (Kamarck et al., 1992; Manuck, 1994), and also reduces the potential for habituation within measurement sessions. However, aggregating across different types of stressors may not improve temporal consistency because different stressors can have different psychological and physiological effects (e.g., Lovallo, 1997), and individual differences in response to one type of stressor may not generalize to individual differences on another (Berntson, Cacioppo, & Fieldstone, 1996). Examination of Table 2 shows that year-to-year Spearman coefficients for aggregated stressor scores were no higher than the highest coefficients based on the speech or math stressor alone.

The third reason protocol design is important in assessing consistency of reactivity is that, depending on the nature of the protocol, participants' perceptions of the stress tasks may vary within or across measurement occasions. Standardized stressors that can be calibrated according to the participants' performance levels (such as the subtraction task used here) help to ensure that the participants all experience approximately the same amount of stress within each measurement occasion and, relative to each other, from occasion to occasion. However, it is also necessary to vary the tasks enough to prevent habituation across the retest interval. In our protocol, we varied the topic of the speech task from year to year. Examination of the significance levels and effect sizes for reactivity in both years, along with the general lack of significant year-to-year declines in mean reactivity (Table 1), suggest that the impact of the stressors diminished very little from Year 1 to Year 2.

In the current study, year-to-year consistency estimates were lower for reactivity scores than for baseline, stressor, or posttask levels for all variables except PEP. Because measurement error is compounded when change scores are created, these results were expected. Nevertheless, there was a wide range of disparity between the reliabilities of the reactivity scores and of their component baseline and stressor scores.

Among the cardiovascular variables, for example, the year-to-year consistency for HR reactivity was only slightly lower than for baseline and stressor scores. This lower consistency of reactivity scores may be due primarily to measurement error. In other words, it appears that an individual's HR reactivity to psychological stress is relatively stable from year to year in this sample. On the other hand, the year-to-year consistency of RSA reactivity was extremely low, in contrast to its values for baseline and stressor. HR is influenced strongly by both sympathetic and parasympathetic input to the heart. Because RSA is an index of cardiac parasympathetic tone, a strong contributor to HR, it is somewhat surprising that consistency for RSA reactivity was so much lower than for HR reactivity in this sample.

The low consistency estimate for RSA reactivity may be due in part to the older age of our sample. Aging is associated with a progressive decrease in vagal control of the heart (Saul & Cohen, 1994), thus vagal reactivity to stress may become less consistent from one episode to another in older individuals. Alternatively, the consistency of RSA reactivity might in fact have been higher than indicated by the consistency coefficients. Because stress most often leads to decreases in vagal tone and hence in RSA, age-related low values at baseline might have reduced the range of change scores and attenuated the year-to-year correlations (see Taylor, Hayano, & Seals, 1995, for a similar hypothesis regarding vagal response to exercise stress). Another potential contributing factor is the consistent use of estrogen replacement by approximately one-third of the sample. Previous work by the current authors documented differences in magnitude of RSA reactivity to stress between women who used estrogen replacement and those who did not (Burleson et al., 1998, 2000); however, no information exists regarding the possible effects of estrogen therapy on the temporal consistency of RSA reactivity.

PEP was the only variable that demonstrated slightly higher consistency for reactivity than for baseline or stressor measures. This occurred despite the fact that stressor-related changes in PEP did not achieve statistical significance in either year. We

speculate that these results may stem in part from the measurement issues described above—namely, difficulty in picking the B-point in older participants. PEP baselines may have shown lower temporal consistency because of the difficulty of consistent B-point placement from year to year, whereas PEP reactivity was more consistent because once the B-point was determined, B-point placement was consistent across each year's baseline and stressor within each participant.

Consistency values for SBP and DBP reactivity fell between the extremes represented by HR and RSA. The estimates for SBP were considerably lower than many of those found previously in several studies using retest intervals longer than one year. Prior estimates for DBP consistency tended to be much lower than for SBP; hence the current results for DBP are less discrepant from previous findings. Although we have no definitive explanation, we speculate that our BP methodology may be partly responsible. We used continuous tonometric recordings made from the radial artery to determine BP. It may be the case that radial tonometry is less able to accurately track changes in BP than auscultatory methods, particularly in older individuals with diminished radial artery distensibility.

Year-to-year consistency coefficients for neuroendocrine change scores were generally low. The values for cortisol were much lower than those for the other three hormones, and also considerably lower than previous estimates ($r = .37$; Cohen et al., 2000; rs = .17–.60, Kirschbaum et al., 1995). Prior investigations of temporal consistency of cortisol reactivity were carried out using saliva rather than peripheral blood. Saliva allows the measurement of free (unbound) cortisol, whereas our assay using plasma measured total cortisol. This difference may in part explain the lower consistency estimate we obtained in the current sample. In addition, cortisol levels did not increase significantly after the stressors in either year, perhaps due to the fact that the measurements were taken during the steep downward portion of the diurnal cycle. This diurnal influence might also reduce the consistency of reactivity values.

No previous estimates exist for temporal consistency of ACTH reactivity; we provide the first. For the catecholamines, prior research has documented low to moderate consistency of reactivity values over time intervals up to 12 weeks in length; the current values support a similar level of consistency in the ranks of reactivity values over a much longer retest interval.

Implications for Future Stress Research

The current study demonstrates that in a sample of mid-aged and older women, statistically significant consistency coefficients can be achieved for reactivity of some stress-related physiological parameters over an interval of approximately 1 year. However, none of the delta scores reached a conventionally "high" level of temporal consistency (i.e., $\geqslant .80$) for any of the variables measured in this sample. On the other hand, many of the baseline, stressor, and posttask values in the current study were highly consistent. What are the implications for researchers studying stress responses? At the very least, these results reinforce the need to minimize measurement error in longitudinal studies of stress reactivity, particularly in older individuals. The use of more frequent measurement and aggregation across measurements, particularly when measuring hormones, should help to achieve this goal.

Using the current protocol, we obtained moderately to highly consistent estimates of individual differences in baseline scores, supporting the usefulness of long-term studies of tonic physiological responses. We also found high consistency in individual differences for many stressor responses and posttask measures, suggesting that researchers should further explore their predictive value when studying physiological responses to psychological challenge. However, we believe that the assessment of reactivity (as well as recovery) is also important for a complete understanding of the effects of stress. Further improvements in the temporal consistency of these measures can likely be made through careful attention to protocol design and analysis strategies. The inclusion of diverse stressors and frequent measures (both for aggregation within measurement occasions and to assess recovery) and the creative design of stress tasks both to enhance and standardize their impact over time are critically important.

REFERENCES

Allen, M. T., Sherwood, A., Obrist, P. A., Cromwell, M. D., & Grange, L. A. (1987). Stability of cardiovascular reactivity to laboratory stressors: A 2 1/2 year follow-up. *Journal of Psychosomatic Research*, *31*, 639–645.

Berntson, G., Cacioppo, J., & Fieldstone, A. (1996). Illusions, arithmetic, and the bidirectional modulation of vagal control of the heart. *Biological Psychology*, *44*, 1–17.

Berntson, G. G., Bigger, J. T. J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & Molen, M. W. v. d. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*, 623–648.

Berntson, G. G., Cacioppo, J. T., Binkley, P. F., Uchino, B. N., Quigley, K. S., & Fieldstone, A. (1994). Autonomic cardiac control. III. Psychological stress and cardiac responses in autonomic space as revealed by automonic blockades. *Psychophysiology*, *31*, 599–608.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, *98*, 459–487.

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

Burleson, M., Wright, R., Baller, M., Farr, C., Cole, J., Mine, L., Kazaka, K., Waldie, C., & Hart, M. (2000, October). *Postmenopausal hormone replacement: Effects on vagal response to psychological versus physiological demand.* Poster presented at the Society for Psychophysiological Research, San Diego, CA.

Burleson, M. H., Malarkey, W. B., Cacioppo, J. T., Poehlmann, K. M., Kiecolt-Glaser, J. K., Berntson, G. G., & Glaser, R. (1998). Postmenopausal hormone replacement: Effects on autonomic, neuroendocrine, and immune reactivity to brief psychological stressors. *Psychosomatic Medicine*, *60*, 17–25.

Burns, J., Ferguson, M., Fernquist, S., & Katkin, E. (1992). Test–retest reliability of inotropic and chronotropic measures of cardiac reactivity. *International Journal of Psychophysiology*, *12*, 165–168.

Byrne, E. A., Slater, B. A., & Porges, S. W. (1991). Respiratory sinus arrhythmia (V) is stable over a 3-week period in adults (Abstract). *Psychophysiology*, *28*, S15.

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

Cacioppo, J. T., Malarkey, W. B., Kiecolt-Glaser, J. K., Uchino, B. N., Sgoutas-Emch, S. A., Sheridan, J. F., Berntson, G. G., & Glaser, R.

(1995). Heterogeneity in neuroendocrine and immune responses to brief psychological stressors as a function of autonomic cardiac activation. *Psychosomatic Medicine*, *57*, 154–164.

Cohen, S., Hamrick, N., Rodriguez, M. S., Feldman, P. J., Rabin, B. S., & Manuck, S. B. (2000). The stability of and intercorrelations among cardiovascular, immune, endocrine, and psychological reactivity. *Annals of Behavioral Medicine*, *22*(3), 171–179.

Coste, J., Strauch, G., LeTrait, M., & Bertagna, S. (1994). Reliability of hormonal levels for assessing the hypothalamic–pituitary–adrenocortical system in clinical pharmacology. *British Journal of Clinical Pharmacology*, *38*, 474–479.

Dimsdale, J. E., & Ziegler, M. G. (1991). What do plasma and urinary measures of catecholamines tell us about human response to stressors? *Circulation*, *83*(suppl. II), 36–42.

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

Goldstein, D. S., Eisenhofer, G., Sax, F. L., Keiser, H. R., & Kopin, I. J. (1987). Plasma norepinephrine pharmacokinetics during mental challenge. *Psychosomatic Medicine*, *49*, 591–605.

Graham, J., Hofer, S., Donaldson, S., MacKinnon, D., & Schafer, J. (1997). Analysis with missing data in prevention research. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–366). Washington, DC: American Psychological Association.

Grassi, G., Bolla, G., Servalle, G., Turri, C., Lanfranchi, A., & Mancia, G. (1997). Comparison between reproducibility and sensitivity of muscle sympathetic nerve traffic and plasma noradrenaline in man. *Clinical Science*, *92*, 285–289.

Hatch, J. P., Borcherding, S., & Norris, L. K. (1990). Cardiopulmonary adjustments during operant heart rate control. *Psychophysiology*, *27*, 641–647.

Hjemdahl, P., Freyschuss, U., Juhlin-Dannfelt, A., & Linde, B. (1984). Differentiated sympathetic activation during mental stress evoked by the Stroop test. *Acta Physiologica Scandinavica Supplement*, *527*, 25–29.

Houtman, I. L. D., & Bakker, F. C. (1991). Stress and coping in lecturing, and the stability of responses across practice. *Journal of Psychosomatic Research*, *35*, 323–333.

Huizenga, N. A. T. M., Koper, J. W., Lange, P. d., Pols, H. A. P., Stolk, R. P., Grobbee, D. E., Jong, F. H. d., & Lamberts, S. W. J. (1998). Interperson variability but intraperson stability of baseline plasma cortisol concentrations, and its relation to feedback sensitivity of the hypothalamic–pituitary–adrenal axis to a low dose of dexamethasone in elderly individuals. *Endocrinology and Metabolism*, *83*, 47–54.

Jern, S., Pilhall, M., Jern, C., & Carlsson, S. G. (1991). Short-term reproducibility of a mental arithmetic stress test. *Clinical Science*, *81*, 593–601.

Kamarck, T., Jennings, J., Debski, T., Glickman-Weiss, E., Johnson, P., Eddy, M., & Manuck, S. (1992). Reliable measures of behaviorally evoked cardiovascular reactivity from a PC-based test battery: Results from student and community samples. *Psychophysiology*, *29*, 17–28.

Kamarck, T. W., Jennings, J. R., Stewart, C. J., & Eddy, M. J. (1993). Reliable responses to a cardiovascular reactivity protocol: A replication study in a biracial female sample. *Psychophysiology*, *30*, 627–634.

Kasprowicz, A. L., Manuck, S. B., Malkoff, S. B., & Krantz, D. S. (1990). Individual differences in behaviorally evoked cardiovascular response: Temporal stability and hemodynamic patterning. *Psychophysiology*, *27*, 605–619.

Kelsey, R., & Guethlein, W. (1990). An evaluation of the ensemble averaged impedance cardiogram. *Psychophysiology*, *27*, 24–33.

Kirschbaum, C., Prussner, J., Stone, A., Federenko, I., Gaab, J., Lintz, D., Schommer, N., & Hellhammer, D. (1995). Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men. *Psychosomatic Medicine*, *57*, 468–474.

Kirschbaum, C., Steyer, R., Eid, M., Patalla, U., Schwenkmezger, P., & Hellhammer, D. H. (1990). Cortisol and behavior: 2. Application of a latent state–trait model to salivary cortisol. *Psychoneuroendocrinology*, *15*(4), 297–307.

Kirschbaum, C., Wust, S., Faig, H.-G., & Hellhammer, D. H. (1992). Heritability of cortisol responses to human corticotropin-releasing hormone, ergometry, and psychological stress in humans. *Journal of Clinical Endocrinology and Metabolism*, *75*, 1526–1530.

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

Kochiadakis, G., Orfanakis, A., Rombola, A., Chrysostomakis, S., Chlouverakis, G., & Vardas, P. (1997). Reproducibility of time-domain indexes of heart rate variability in patients with vasovagal syncope. *American Journal of Cardiology*, *79*, 160–165.

Krantz, D. S., & Manuck, S. B. (1984). Acute psychophysiologic reactivity and risk of cardiovascular disease: A review and methodologic critique. *Psychological Bulletin*, *96*, 435–464.

Linden, W., Earle, T. L., Gerin, W., & Christenfeld, N. (1997). Physiological stress reactivity and recovery: Conceptual siblings separated at birth? *Journal of Psychosomatic Research*, *42*(2), 117–135.

Linden, W., Rutledge, T., & Con, A. (1998). A case for the usefulness of laboratory social stressors. *Annals of Behavioral Medicine*, *20*(4), 310–316.

Llabre, M. M., Saab, P. G., Hurwitz, B. E., Schneiderman, N., Frame, C. A., Spitzer, S., & Phillips, D. (1993). The stability of cardiovascular parameters under different behavioral challenges: One-year follow-up. *International Journal of Psychophysiology*, *14*, 241–248.

Lovallo, W. (1997). *Stress and health: Biological and psychological interactions*. Thousand Oaks, CA: Sage Publications.

Manuck, S. B. (1994). Cardiovascular reactivity in cardiovascular disease: "Once more unto the breach". *International Journal of Behavioral Medicine*, *1*, 4–31.

Manuck, S. B., Kamarck, T. W., Kasprowicz, A. S., & Waldstein, S. R. (1993). Stability and patterning of behaviorally evoked cardiovascular reactivity. In James J. Blascovich, & Edward S. Katkin (Eds.). *Cardiovascular reactivity to psychological stress & disease* (pp. 111–134). Washington, DC: American Psychological Association.

Manuck, S. B., Kasprowicz, A. L., & Muldoon, M. F. (1990). Behaviorally-evoked cardiovascular reactivity and hypertension: Conceptual issues and potential associations. *Annals of Behavioral Medicine*, *12*, 17–29.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *The New England Journal of Medicine*, *338*(3), 171–179.

Mills, P. J., Berry, C. C., Dimsdale, J. E., Neleson, R. A., & Ziegler, M. G. (1993). Temporal stability of task-induced cardiovascular, adrenergic, and psychological responses: The effects of race and hypertension. *Psychophysiology*, *30*, 197–204.

Mills, P. J., Ziegler, M. G., Dimsdale, J. E., & Parry, B. L. (1995). Enumerative immune changes following acute stress: Effect of the menstrual cycle. *Brain, Behavior, and Immunity*, *9*, 190–195.

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

Porges, S., & Bohrer, R. (1990). The analysis of periodic processes in psychophysiological research. In J. Cacioppo, & L. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 708–753). Cambridge: Campridge University Press.

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., Auer, K. v., Jobst, S., Kaspers, F., & Kirschbaum, C. (1997). Free cortisol levels after awakening: A reliable biological marker for the assessment of adrenocortical activity. *Life Sciences*, *61*, 2539–2549.

Rose, R. (1984). Overview of endocrinology of stress. In G. M. Brown, S. H. Koslow, & S. Reichlin, (Eds.), *Neuroendocrinology and psychiatric disorder* (pp. 95–122). New York: Raven Press.

Saab, P. G., Matthews, K. A., Stoney, C. M., & McDonald, R. J. (1989). Premenopausal and postmenopausal women differ in their cardiovascular and neuroendocrine responses to behavioral stressors. *Psychophysiology*, *26*, 270–280.

Saul, J. P., & Cohen, R. J. (1994). Respiratory sinus arrhythmia. In M. N. Levy, & P. J. Schwartz (Eds.), *Vagal control of the heart: Experimental basis and clinical implications* (pp. 511–536). Armonk, NY: Futura Publishing.

Schulz, P., & Knabe, R. (1994). Biological uniqueness and the definition of normality. Part 2—The endocrine "fingerprint" of healthy adults. *Medical Hypotheses*, *42*, 63–68.

Seeman, T. E., & Robbins, R. J. (1994). Aging and hypothalamic–pituitary–adrenal response to challenge in humans. *Endocrine Reviews*, *15*(2), 233–260.

Seraganian, P., Hanley, J. A., Hollander, B. J., Roskies, E., Smilga, C., Martin, N. D., Collu, R., & Oseasohn, R. (1985). Exaggerated psychophysiological reactivity: Issues in quantification and reliability. *Journal of Psychosomatic Research*, *29*, 393–405.

Sherwood, A., Allen, M., Fahrenberg, J., Kelsey, R., Lovallo, W., & Dooren, L. v. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, *27*, 1–23.

Sherwood, A., Girdler, S. S., Bragdon, E. E., West, S. G., Brownley, K. A., Hinderliter, A. L., & Light, K. C. (1997). Ten-year stability of cardiovascular responses to laboratory stressors. *Psychophysiology*, *34*, 185–191.

Sherwood, A., Turner, J. R., Light, K. C., & Blumenthal, J. A. (1990). Temporal stability of the hemodynamics of cardiovascular reactivity. *International Journal of Psychophysiology*, *10*, 95–98.

Siever, L. J., Uhde, T. W., Jimerson, D. c., Lake, C. R., Kopin, I. J., & Murphy, D. L. (1986). Indices of noradrenergic output in depression. *Psychiatric Research*, *19*, 59–73.

Sloan, R. P., Shapiro, P. A., Bagiella, E., Gorman, J. M., & Bigger, J. T. (1995). Temporal stability of heart period variability during a resting baseline and in response to psychological challenge. *Psychophysiology*, *32*, 191–196.

Smyth, J., Ockenfels, M. C., Gorin, A. A., Catley, D., Porter, L. S., Kirschbaum, C., Hellhammer, D. H., & Stone, A. A. (1997). Individual differences in the diurnal cycle of cortisol. *Psychoneuroendocrinology*, *22*(2), 89–105.

Stewart, J., & France, C. (2001). Cardiovascular recovery from stress predicts longitudinal changes in blood pressure. *Biological Psychology*, *58*(2), 105–120.

Stoney, C., Niaura, R., & Bausserman, L. (1997). Temporal stability of lipid responses to acute psychological stress in middle-aged men. *Psychophysiology*, *34*(3), 285–291.

Stoney, C. M., & Hughes, J. W. (2001). Catecholamine stress responses in arterialized blood. *Psychophysiology*, *38*, 590–593.

Strube, M. J. (1990). Psychometric principles: From physiological data to psychological constructs. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 34–57). New York: Cambridge University Press.

Swain, A., & Suls, J. (1996). Reproducibility of blood pressure and heart rate reactivity: A meta-analysis. *Psychophysiology*, *33*, 162–174.

Taylor, J. A., Hayano, J., & Seals, D. R. (1995). Lesser vagal withdrawal during isometric exercise with age. *Journal of Applied Physiology*, *79*(3), 805–811.

Thomas, G., Frenoy, N., Legrain, S., Sebag-Lanoe, R., Baulieu, E.-E., & DeBuire, B. (1994). Serum dehydroepiandrosterone sulfate levels as an individual marker. *Journal of Clinical Endocrinology and Metabolism*, *79*, 1273–1276.

Uchino, B. N., Kiecolt-Glaser, J. K., & Cacioppo, J. T. (1992). Age-related changes in cardiovascular response as a function of a chronic stressor and social support. *Journal of Personality and Social Psychology*, *63*, 839–846.

van de Borne, P., Montano, N., Zimmerman, B., Pagani, M., & Somers, V. (1997). Relationship between repeated measures of hemodynamics, muscle sympathetic nerve activity, and their spectral oscillations. *Circulation*, *96*, 4326–4332.

Veith, R., Best, J., & Halter, J. (1984). Dose-dependent suppression of norepinephrine appearance rate in plasma by clonidine in man. *Journal of Clinical Endocrinology and Metabolism*, *59*, 151–155.

Willemsen, G., Ring, C., Carroll, D., Evans, P., Clow, A., & Hucklebridge, F. (1998). Secretory immunoglobulin A and cardiovascular reactions to mental arithmetic and cold pressor. *Psychophysiology*, *35*, 252–259.

Young, E., Aggen, S., Prescott, C., & Kendler, K. (2000). Similarity in saliva cortisol measures in monozygotic twins and the influence of past major depression. *Biological Psychiatry*, *48*, 70–74.

Ziegler, M. G. (1989). Catecholamine measurement in behavioral research. In N. Schneiderman, S. M. Weiss, & P. G. Kaufmann (Eds.), *Handbook of research methods in cardiovascular behavioral medicine* (pp. 167–184). New York: Plenum Press.

Zorilla, E. P., DeRubeis, R. J., & Redei, E. (1995). High self-esteem, hardiness and affective stability are associated with higher basal pituitary–adrenal hormone levels. *Psychoneuroendocrinology*, *20*, 591–601.

(Received September 27, 2000; Accepted October 22, 2002)