A CASE FOR THE USEFULNESS OF LABORATORY SOCIAL STRESSORS[1,2]

Wolfgang Linden, Thomas Rutledge, and Andrea Con
The University of British Columbia

ABSTRACT

Although laboratory stress research is a popular and vibrant area of research activity, there is surprisingly little evidence that laboratory stress models are clinically useful (i.e. that they can explain and predict the development of disease). This article summarizes evidence that the usefulness of lab stress research can be improved with the use of social stressors. Two lines of evidence are presented in support of this argument: (a) studies comparing physiological reactivity to different lab stressors with ambulatory activity, and (b) a meta-analysis of investigations of cortisol responses to laboratory stressors. Further issues of importance in understanding social stressors are gender differences and the vulnerability (i.e. weak reliability) of social stressor impact to relatively small changes in the experimental protocol itself.

(**Ann Behav Med** 1998, *20*(4):310–316)

INTRODUCTION

Can the study of acute cardiovascular reactivity in the lab teach us about the development of disease? This question about the usefulness of reactivity research has received answers that vary from a very conservative and well-argued conclusion that "there is very little to suggest it" (1) to a more optimistic endorsement that, however, still points out the many critical gaps in the available literature (2).

In order to illustrate the many ways in which usefulness can be shown, typical research strategies involving lab stressors will be sketched out and a case will be made for the superior usefulness of social stressors. This case for social stressors will be supported by data from two particular lines of evidence: (a) via improving the predictive value of lab resting and reactivity indices for the more clinically relevant ambulatory blood pressure and (b) by showing distinct cortisol responses to social stressors.

In essence, cardiovascular lab stress testing is a psychological test and needs to meet the same quality criteria as other tests: that is, reliability and validity need to be shown. This article is particularly concerned with issues of response stability over time and across situations (reliability) and with predictive and explanatory usefulness (i.e. predictive and criterion validity).

[1] Preparation of this manuscript was supported in part by financial support awarded to the first author by the Social Sciences and Humanities Research Council of Canada as well as the British Columbia and Yukon Heart and Stroke Foundation.

[2] This paper is based on symposia presentations at the Annual Meetings of the American Psychological Association, Toronto, Canada, August 1996 and of the Society for Psychophysiological Research, Vancouver, Canada, October 1996.

Reprint Address: W. Linden, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver B.C., V6T 1Z4, Canada.

EVIDENCE OF STABILITY

Most studies on the stability of reactivity indices suggest small to moderate test–retest correlations (cf. 3), although recent evidence suggests that with refined protocols test–retest correlation coefficients can approach or even surpass the .8 mark (4,5). One particular protocol feature accounting for the greater gain in stability in the Kamarck et al. (4) study was the computation of an aggregate (or composite) score derived by averaging the observed changes from three tasks into a single reactivity index. The greater stability gained by computing a composite score is not surprising given that each physiological measurement point is equivalent to a test item on a personality measure and that test reliability increases as the number of measures (or test items) increase provided they themselves are intercorrelated (4). On a qualitative level, it is also likely that a larger sample of reactivity indices is more representative of real-life physiological activity (as measured with ambulatory monitors) and therefore easier to reproduce. Because Kamarck et al. (4) used a series of lab-based active coping tasks, their findings speak to response stability for this class of tasks; but reactivity indices across different classes of tasks are much lower, suggesting little overlap in explained variance (6,7).

The discrepancy in observed response stability between repetitions of similar active coping tasks versus the presentation of a more diverse test package requires further analysis and interpretation. It is well-established that different tasks are also associated with differential cardiovascular response patterns, and these, in turn, may reflect qualitative differences in emotional response, varying intensities of cognitive challenge and needed effort at resolution, and distinct underlying hemodynamic response patterns. One such response distinction is that of an alpha- versus a beta-adrenergic pattern, with a more predominant vascular resistance component in the former (indexed by arterial constriction, smooth muscle contraction, inhibition of renin release) and a pronounced cardiac, sympathetic component in the beta-adrenergic response (indexed by heart rate and rate-pressure product increase, renin release, and skeletal muscle contraction) (8,9).

The differences in emotional quality triggered by different tasks are poorly understood (10) and have so far not been very revealing as self-report indices of affect tend to correlate poorly with physiological changes.

ON THE UTILITY OF STRESS REACTIVITY RESEARCH

Different strategies for determining the utility of stress reactivity research can be gleaned from a selective review of relevant studies; some involve physiological measurement, others do not.

One line of evidence supporting the use of social stressors is derived from diary studies of mood changes as a function of type of preceding stressor (11). DeLongis and collaborators found that their subjects recovered quickly in their self-rated moods following a variety of different stressors with, however, a noticeable exception in that interpersonal stressors were associated with very

slow mood recovery (i.e. extended reports of psychological distress).

Also, because hostility/anger are considered important emotional variables for the development of cardiovascular disease (CVD), studies of anger/hostility in the lab can provide important clues. Suls and Wan (12) conducted a meta-analysis of lab stress studies that targetted the link between anger hostility as a trait and a variety of lab stressors, some of which were specifically designed to trigger anger. These researchers showed that hostility was only associated with cardiovascular hyperresponding when the stressor itself was interpersonally harrassing and provocative in nature. This usually involved the direct interaction of the experimental subject with either a trained experimenter or confederate who was instructed to be provocative. Such provocation might consist of unfair criticism of someone's performance or a pointedly worded instruction to perform more quickly.

Next, one can compare response magnitudes associated with different challenges on the presumption that tasks triggering greater responses may be more relevant for disease development as "vulnerability" theories suggest (cf. 2,13). Although interpersonal stressors have often been shown to produce greater physiological responses (especially alpha-adrenergic responses) than more cognitive tasks (7,14–16), there is as yet no evidence that one class of task is a better predictor of disease outcome than another type of task. Indeed, the entire reactivity construct has not been clearly shown to directly predict coronary heart disease (CHD) mortality or morbidity (1), although it is widely believed to contribute to the etiology of CHD (2,17).

Because prospective studies are difficult, expensive, and logistically challenging, a more intermediate test of usefulness is to relate lab reactivity to ambulatory blood pressure studies. Theoretically, one could show usefulness of the lab stress reactivity construct by demonstrating that knowledge of lab reactivity can permit better prediction of blood pressure change over time than is possible with the knowledge of resting values alone (cf. 15). This approach to testing usefulness is a specific focus of this paper.

There is suggestive evidence that lab responses are predictive of long-term changes in blood pressure (BP) (see 2, 13, or 15 for reviews), and elevated blood pressure in turn is a well-established major risk factor for CHD and stroke. Barnett, Spence, Manuck, and Jennings (18) support the usefulness of the reactivity concept by showing that lab hyperresponders showed more rapid progression of atherosclerotic plaques over a two-year observation period than did a less responsive cohort. There also is evidence from a five-year follow-up of young borderline hypertensives that slow recovery from lab stress induced BP increases is an important negative prognostic indicator for hypertension development (19). In this study, slow recovery from a cognitive challenge was a more useful predictor of long-term BP changes than was initial reactivity (i.e. baseline to task change).

STUDIES OF AMBULATORY BP AND LAB REACTIVITY LINKAGE

The use of ambulatory pressures as an intermediary clinical endpoint is well-justified: 24-hour averaged ambulatory blood pressures are better long-term predictors of the development of hypertension than lab resting measures (20,21), averaged ambulatory blood pressures relate more closely to target-organ damage than do lab measures (22,23), and drug treatment research increasingly uses ambulatory measures as clinical endpoints. While costly and cumbersome to obtain, ambulatory-derived averages appear to have become the "gold standard" for blood pressure measurement. In consequence, the usefulness of the lab reactivity paradigm would be strengthened if consistency (i.e. concurrent validity) between lab measurement models and ambulatory blood pressures could be demonstrated. A number of researchers have reported consistency data, but the intercorrelations fall mostly in the .2 to .6 range, thus explaining only small portions of the variance in ambulatory means (cf. 24–28). Hence, the question arises whether refined methods and protocols can lead to greater consistency between blood pressures measured in the two environments. One such strategy would be to improve the verisimilitude (or real-life equivalence) of lab stress; this is what the use of social stressors is all about. Lay people and researchers are generally easy to convince that social stressors, like a marital dispute or an assertive challenge, possess content validity for real life, and this was further supported by the mood and stress diary study described earlier (11).

At this time, a number of studies on lab–ambulatory linkage are available, and their protocols can be subdivided into two main types. One type of protocol is predominantly concerned with overlap in real-life reactivity and lab reactivity to standardized stressors (29,30). These authors noted that a within-subject design was necessary to demonstrate consistency between lab reactivity and ambulatory blood pressure reactivity to everyday stressors. Few studies of this type have been conducted to date, largely because concepts like a baseline or a recovery period can not be readily defined in the real world. The resulting data analyses and interpretations are very difficult and ambiguous.

A second type of protocol ties lab reactivity protocols to mean ambulatory pressures. In this type of study, researchers can first determine whether lab resting measures (singular or aggregated) are strong predictors of ambulatory means. Next, even if reactivity in the lab is predictive of ambulatory cardiovascular activity, it remains to be determined whether or not the study of reactivity to laboratory challenge adds to existing predictor models that use lab resting values as predictors of ambulatory means. This is important because reactivity studies are considerably more time-consuming and expensive than is the determination of resting measures, and it would be redundant to invest in the study of reactivity if it does not uniquely contribute information to understanding the etiology and prediction of hypertension (or any other poor health outcome).

We know of five studies that used comparable lab/real-world protocols with multitask presentations in the lab, and each of these studies had explicitly tested whether addition of the reactivity information improved the predictor model for lab versus ambulatory measurement.

Langewitz et al. (27) noted that the investigation of reactivity to math or the cold pressor did not provide a better prediction than did resting values as predictors. Subjects were 77 male hypertensive patients, but in some of the comparisons only about half of the sample's data could be used.

In a study by Ironson, Gellman, Spitzer, et al. (26), 119 men and women were monitored during four lab challenges (structured interview, video game, bicycle exercise, and cold pressor test) and also with ambulatory BP monitors during 14 waking hours (9 a.m. to 11 p.m.), allowing them to further differentiate work and home levels of cardiovascular activity. When averaged across male and female, Black and White subjects, only the response to the structured interview added additional explained variance to the baseline ambulatory predictor model (a 3% gain in explained variance was found). Interestingly, the reactivity to the cold pressor was a superior predictor to the interview response in Blacks, whereas the opposite was true for Whites.

TABLE 1
Summary Data from Studies Using Lab Predictors of Ambulatory Blood Pressure Means

Study	Type of Task		
	Physical Effort	Cognitive	Interpersonal
Langewitz et al. (27)	n/a	−	n/a
Ironson et al. (26)	−	−	+
Ewart and Kolodner (31)	n/a	−	+
Linden and Con (32)	−	−	+
Light et al. (15)	n/a	−	+

Notes: n/a = the study had no such task; − = the task response did not improve the prediction model; + = the task response did improve the prediction model.

Similarly, Ewart and Kolodner (31) noted that a social interaction task (the Social Competence Interview) contributed more useful information for predicting ambulatory pressures in adolescents than did other active coping tasks (mirror drawing, mental arithmetic, video game). Original to this study was the use of an overall lab average, where all available information from adaptation, baseline, task, and recovery phases was pooled into a single score. This overall lab mean was a noticeably better predictor of ambulatory systolic blood pressure (SBP) means than either lab resting values or any combination of lab resting and reactivity indices. Although not significant, a similar trend suggesting the usefulness of an overall lab mean score was apparent for studying the lab–ambulatory linkage of diastolic blood pressure (DBP) and heart rate (HR) measures.

In a sample of 148 White and Black women and men, Light et al. (15) replicated the race-related differences in lab/real-life prediction. The tasks used in the lab were a computerized math challenge, a competitive reaction time task, a frustrating role-play about hassles with a vendor, listening to a competitor's speech, and the forehead cold pressor task. Cold pressor response again improved the lab resting values prediction model for Blacks, whereas an active speech task was the best predictor for Whites. Also, when pooled across genders and races, the active speech task was the relatively most useful contributor to explaining variance in ambulatory BP in addition to resting values.

Our own group (32) conducted a similar study with 126 university students, an 8- to 10-hour ambulatory monitoring period on campus, and three lab tasks: mental arithmetic, discussion of the emotional response to a past frustrating event, and a handgrip task. The findings were much like those of previous researchers in that only the reactivity to the discussion task enhanced the baseline ambulatory prediction model, and only on DBP. Similar to Ewart and Kolodner's findings (31), the aggregation of all SBP scores from the entire lab portion of the study enhanced the prediction of ambulatory SBP, whereas no single reactivity index improved the SBP lab to real-life prediction. Together, these studies served as partial replications to one another and revealed a fairly consistent pattern of results that is summarized in Table 1.

REACTIVITY VERSUS AMBULATORY STUDIES

The studies described above are sufficiently similar to allow comparison but also different enough (racial compositions of samples, some variation in task choices, and different aged samples) so as to permit good generalizability. The pattern of results appears quite consistent. Lab resting values of BP are already strong predictors of ambulatory BP; aggregation of scores in the lab tend to improve the predictor model (typically more so on systolic blood pressure), whereas reactivity tends to add little to the predictor model. The relatively most useful reactivity index appears to be social task responses which (despite varying protocols) showed repeated advantages over other reactivity indices. This observation needs to be moderated, however, in that speech appears to improve the predictor models for Whites but not necessarily for Blacks, and it may be strongest for diastolic BP (i.e. a vasoconstrictive index).

These observations in and of themselves allow no definitive explanation as to why an overall lab mean score or the addition of psychosocial stressors may be useful. One plausible argument is that when more measures are available from every individual, it simply creates a more reliable test (see Kamarck's [4] suggestion of composite or aggregate measures). A second explanation is theoretically compatible with the aggregation argument and also centers on representativeness; it differs, however, in that the additional study of adaptation and recovery phases not only increases test reliability, but also samples activities and psychological sets (or qualities) that are not represented in the baseline task paradigm. Hence, reliability and validity may be enhanced with multiple measures, a wide range of qualitatively different challenges, and the inclusion of recovery indices. A case in point for the importance of recovery phenomena is the study of the noticeably slow blood pressure and heart rate recovery from anger provocation relative to other challenging tasks which, however, do not trigger anger (33,34) (for a more in-depth review of recovery issues, see 35). In either case, the findings suggest that the use of resting values (in lab or office) as a predictor for the more clinically relevant ambulatory BP can be improved by aggregating across multiple task responses and recovery periods, and the use of social stressors may further improve the predictor model.

Even this preceding discussion does not provide a full explanation of why social interaction tasks may have validity advantages. The discussion task used by Linden and Con (32) was also associated with the greatest blood pressure response compared to the other tasks, and this is a replicated finding (7,29). The propensity of this task to account for more variance in real-life blood pressure (at least for DBP) may simply be related to the fact that the discussion triggered the greatest magnitude of change (or variability) in the lab. One can also argue that the task content of social stressors bears more resemblance to real-life challenges (i.e. possesses content and ecological validity) than does an arithmetic test, a handgrip, or a cold pressor test (10). This is certainly consistent with the DeLongis et al. (11) diary study of mood changes as a function of type of stressor. Finally, it could be that task-specific, physiological response patterns are important. The discussion task in the Linden and Con study (32), for example, was characterized by a greater DBP and a lesser HR response than the math task, suggesting more pronounced alpha-adrenergic activation in the discussion and stronger beta-adrenergic responding during arithmetic performance. It may be that lab predictors capturing alpha-adrenergic (rather than beta-adrenergic) responses have an advantage in predicting ambulatory activity, and social tasks appear to trigger greater alpha-adrenergic responses than do other cognitive challenges. However, this remains speculative since receptor activity is not usually tested in studies with human subjects.

CORTISOL STUDIES

Our second line of argument for social tasks stems from a review of the literature on cortisol responding to acute stressors

(36). Because initial analyses were only published as an abstract, detailed information is not readily available to the reader. For this reason, the data presented as a conference paper in 1994 will be described in more detail below.

The psychophysiological investigation of stress and stress-related diseases is "bread and butter" research for psychophysiologists and has spanned many decades of research efforts. Most investigators focus on a particular disease and/or subsection of physiological systems, like cortical activation, cardiovascular reactivity, or hormonal responses to stress. In the cardiovascular arena, there is growing dissatisfaction with the use of only peripheral measures like blood pressure and heart rate, and as a consequence, some investigators have called for more in-depth examinations of multiple response systems (cf. 37) and for better integration of theoretical models that are used by researchers interested in cardiovascular versus hormonal responsivity (37,38).

Frankenhaeuser (38) has provided a biopsychosocial framework that describes the interaction between peripheral and hormonal indices of stress and distinguishable environmental demands. When a person perceives a stressor, two pathways of physiologic response are activated: the hypothalalmic–pituitary (HP) axis and the sympathetic adrenal–medullary (SA) axis. Psychological, affective distress is primarily reflected in the HP activation system. Motor and cognitive efforts to control a stressor are primarily indexed by catecholamine changes that, in turn, reflect activation of the SA system. SA activation is indexed by a variety of measures with blood pressure, heart rate, and electrodermal activity being frequently-used candidate indices. HP activation results in excessive production of glucocorticoids of which cortisol is a representative index. Major reviews on the topic of stress physiology suggest that activation of the SA axis is not likely to have negative long-term health consequences unless it is also accompanied by activation of the HP system. Furthermore, negative health outcomes do not even require a great deal of SA activation but can be triggered by excessive stimulation of the HP axis alone (37,38). Therefore, measurement of HP activation is highly recommended in stress research, and the recent availability of salivary cortisol sampling has made the addition of a cortisol measure much easier in logistical terms. While the drawing of plasma via in-dwelling catheters typically required a hospital environment and specially-trained staff (thus making research participation rather unattractive to potential subjects), the collection of saliva is simpler for the researcher and less aversive for the subject.

Comparison studies of plasma versus salivary cortisol sampling have suggested overall high comparability of obtained cortisol levels with correlation coefficients ranging from $r = .54$ to $r = .97$ (39). For thorough reviews of the different sampling methodologies and their potential pitfalls, the reader is referred to Kirschbaum and Hellhammer's (39) excellent review. Because the review presented here is concerned with the collection of cortisol data in short-term laboratory challenges, urinary sampling methods are not discussed because the urinary sampling procedure is insensitive to short-term changes typically seen in lab stress experiments.

Given the importance of measuring HP activation in stress research, we were struck by the fact that most of the cortisol research was conducted by investigators focusing on hormonal stress responses only and that cardiovascular researchers (there are hundreds of published studies on cardiovascular reactivity) rarely measured cortical–pituitary activation. The review of the literature (via computer searches and follow-up of secondary references) quickly revealed a paucity of controlled studies that were furthermore difficult to integrate because of an extraordinary variability in type of task used; population sampled; and non-homogeneity of sampling protocols, in particular, the timing of cortisol sampling.

Nevertheless, there appeared to be a reasonable number of studies ($N = 26$) that used lab stressors and that could be subdivided into standardized active coping efforts while alone in the lab (i.e. mental arithmetic, mirror tracing, etc.) versus active coping tasks involving a dyadic interaction and/or performances in front of an audience. Distinguishing tasks on this basis are supported by the observation of generally greater cardiovascular responses for the social interaction over the solo performance studies (cf. 7,16). Some studies provided extensive analysis of within-subject changes in cortisol such that conclusions can be drawn about ideal experimental protocols regarding the necessary length of observation and choice of sampling time points.

In this light, the objectives of the review were reduced to two: (a) compare effect sizes for lab stressors that either involved solo performance tasks or tasks with a social interaction; and (b) uncover the response curves for cortisol over poststressor periods of up to one hour. For the first objective, it was hypothesized that lab challenges with an audience would trigger greater HP activation (as indexed by cortisol) than performance tasks performed alone. The second objective was more exploratory, hopefully revealing a clear pattern of cortisol responding over time in order to facilitate the establishment of sampling protocols with maximal sensitivity and minimal cost.

Method

Pertinent studies were identified via a computer search (34,40–51). The resulting number of references is less than the number of studies because many authors had more than one experimental condition which was then treated as a separate study. Also, many of the cortisol reactivity studies described here had involved studying the effect of caffeine, nicotine, or alcohol and only those conditions in which subjects did not ingest caffeine, nicotine, etc. were used. Furthermore, in order to deal with the heterogeneity of sampling protocols (anything from a single measure taken immediately posttask to 7-, 10-, 12-, and 15-minute intervals extending up to 2 hours were noted), all measures were pooled into three time blocks (or epochs) of 0–19 minutes poststressor, 20–39 minutes poststressor, and 40–60 minutes poststressor. While admittedly coarse, this pooling approach was designed to increase reliability by exploiting the repeated measurement advantage. There were not sufficient numbers of studies to pool beyond 60 minutes of poststressor sampling.

Because various researchers used different units of measurement (mmol/dl, ng/dl) and varying statistics, all raw data were transformed into effect sizes (ES) representing standard deviations of change (formula $ES = [M1 - M2]/SD$) where M1 refers to the raw group mean determined at varying times poststressor, M2 refers to raw group means at the designated baseline, and SD refers to the standard deviation at the baseline. Data were culled and estimated from graphs when tables with raw data were missing; this was done by maximizing the graph size via photocopy "blow up" and by insertion of gridlines. Standard errors of the mean (SEM) as an index of variability were converted to SD where necessary. ES were weighted for sample size. For the interpretation of effect sizes, Cohen's (52) definitions were used and effects of .2 were called small, .5 medium, and .8 a large effect.

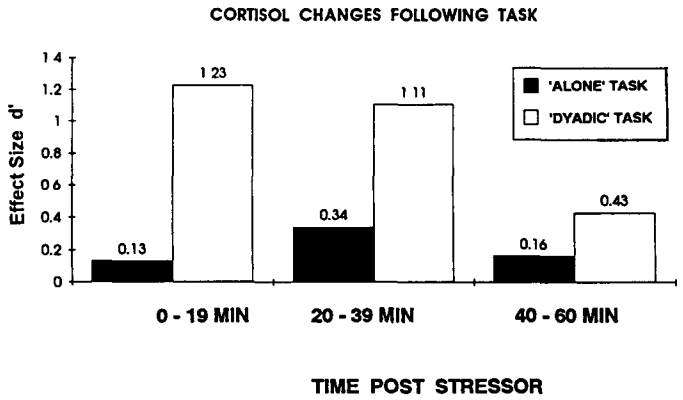

FIGURE 1: Cortisol changes during lab and following lab challenges (expressed as effect size d').

Results

Effect sizes were computed separately for the "solo" tasks versus those tasks that involved a social interaction. Effect sizes were also computed for each reported value in the listed studies but when more than one value per measurement epoch (0–19, 20–39, 40–60 minutes poststressor) was available, all values within one epoch were averaged to permit a standardized mode of analysis and display. The resulting effect sizes for the two different types of tasks and the three successive time epochs are displayed in Figure 1.

Because the number of studies contributing to the computed effect sizes for each time epoch varied, statistical analyses were executed separately for each epoch, contrasting the effect sizes for the reactivity to the classes of tasks.

As the effect sizes displayed in Figure 1 suggest, the social interaction tasks were associated with a much greater cortisol response during the first two epochs (0–19 and 20–39) and a somewhat greater response during epoch 3 (40–60); d = .13 and d = 1.23; d = .34 and d = 1.11; and for epoch 3, d = .16 and d = .43. Effect sizes averaged across the entire 60-minute recovery periods (i.e. averaging across epochs 1, 2, and 3) revealed significantly greater effects associated with the social task (d = .92) than the "alone" tasks (d = .21, t = 2.96, p = .01).

Discussion

Given the small number of available studies, the results should be considered exploratory rather than definitive. Nevertheless, conclusions about when to sample and about task-specific differences in reactivity are apparent even to the naked eye and are confirmed by inferential statistics. Tasks that were interpersonal in nature reliably triggered large cortisol responses that are already apparent at the end of task completion, remain high for about 30–40 minutes, and then decline. Even at the end of a one-hour observation period, the cortisol response still reflects a moderately-sized effect. Tasks performed alone were consistently associated with small cortisol responses without a major decreasing trend over the one-hour observation period. The addition of an audience of some sort significantly raised the cortisol response, averaging an approximate 1:4 ratio in effect sizes. The difference in response magnitude between the two types of tasks cannot be explained by longer exposure to the social stress tasks, because the average task length for the solo was actually longer than for the social tasks (29 versus 20 minutes). This between-task difference in cortisol responding replicates a pattern of findings from cardiovascular reactivity studies where it has also been shown that the involvement of an interviewer (7,32,53) increases the size of the cardiovascular response, with diastolic blood pressure being a particular sensitive index for task differentiation. These findings jointly support Frankenhaeuser's (38) proposition of enhanced distress in challenging social interactions. The large ratio of the difference in cortisol responding to the alone versus social tasks suggests that the addition of a social component greatly affects the HP activation system.

In terms of cortisol sampling points, it appears that Kirschbaum and Hellhammer's (39) suggestion to sample cortisol for at least one hour poststressor is useful and still conservative. Even if cortisol was sampled only for 10–20 minutes poststressor, this period is likely sufficient to detect cortisol responses to potent lab stressors. If, however, researchers are interested in recovery phenomena (and we suggest that they should be), they may still want to sample for at least one hour poststressor.

CONCLUSION

We believe we have built a strong case for the use of interpersonal stressors in reactivity research by using multiple testing strategies. This is, however, not meant to discourage the reader from studying other interesting (often unresolved) issues in reactivity research. First of all, we still need direct supportive evidence that cardiovascular hyperresponses cause cardiac morbidity and mortality. There also is no inherent claim that only interpersonal stressors can be of use. The work on progression of atherosclerosis by Spence et al., for example, produced novel and important results but did not use tasks that we would classify as interpersonal. Also, important links between cardiovascular hyperresponding and immune suppression has been shown by Uchino et al. (54) where the hyperresponse is based on tasks that we also would not classify as interpersonal.

The astute reader will have noticed that we have not provided a definition of social stressors other than having described interpersonal stress protocols. The obvious distinction between interpersonal and non-interpersonal stressors is the presence/absence of another person with whom to directly interact, this being the specific defining ingredient. This does, however, not imply that social elements can not play important roles in seemingly non-interpersonal tasks like math performance. For example, it makes little sense to expect a large cardiovascular response to mental arithmetic unless it is somewhat embarrassing for a subject to perform poorly. Experimenters may still have presence in the minds of the subject who knows that the experimenter will later count correct answers or may be doing so online while watching and listening through one-way mirrors and speaker systems. The subjectively perceived need to do well on math tasks, for example, is acquired via social means: parental or teacher instruction, reinforcement, and modeling. Also, experimenters need to provide instruction or feedback about a task and this makes any task at least somewhat of a social interaction. The potency of lab challenges like Stroop color-word tests, arithmetic, or star-tracing cannot be assumed to lie exclusively in the cognitive performance demands and its associated physiological activation. Fear of negative evaluation, performance anxiety, and a presumed inherent ego-threat are all shared elements of even those tasks labelled non-social in this paper. Therefore, the labelling of tasks as interpersonal versus cognitive does not reflect an inherent, clean dichotomy; it is a question of gradation, of emphasis.

Curiously, although different types of tasks are fairly consistently associated with distinct hemodynamic response patterns (cf. 8), it is much less clear which emotional responses they trigger in experimental subjects. Self-report is typically able to distinguish

high and low stress tasks or high and low anxiety- or anger-provoking tasks, but there is no evidence that any task triggers a pure emotion or provokes the exact same quality of response in all subjects. Habitual styles of subjects to downplay or exaggerate the stressfulness of tasks further muddles the results obtainable from self-report. Experience has repeatedly shown that even using a standardized task like mental arithmetic, we find subjects who rate a given task a "10" on perceived stressfulness whereas another subject rates it a "1."

Throughout this paper we have reported differences in lab stress reactivity between races and there also is evidence for gender differences. For example, in our studies on anger provocation (33,34), we found that women show marginally smaller SBP and HR responses to anger provocation than do men, but women's DBP responses were less than half the effect size of men's responses. This difference was further supported by a much smaller cortisol response to anger in women (34), whereas men and women did not differ much in response magnitudes to arithmetic challenges without harassment.

One final observation regarding the use of interpersonal stressors relates to a relative "fickleness" of interpersonal stressors in that social psychologists have previously shown that relatively small variations in protocol (see, for example, Asch's early studies on interpersonal influence [55] or Gerin et al.'s work on social support [56]) show large between-condition differences in behavior and physiological response. Another example is that even the repeated use of the same harassment challenge in the same laboratory using an identical script (33 versus 34) resulted in smaller mean effects on HR and DBP in the second study (57). In the absence of an obvious explanation, we speculated that repeated use of a powerful harassment manipulation may result in growing sensitivity among experimenters about this method, thereby reducing experimenter comfort and ultimately the potency of this manipulation.

In sum, we presented evidence for the usefulness of social lab stressors. Continued use of social stressors may be maximally useful if researchers pay close attention to subtle protocol variations, and a thorough knowledge of social psychology and its research base is posited to greatly facilitate the design of clean and reliable social stressor protocols.

REFERENCES

(1) Pickering TG, Gerin W: Blood pressure reactivity: Cardiovascular reactivity in the laboratory and the role of behavioral factors in hypertension: A critical review. *Annals of Behavioral Medicine.* 1990, *12:*3–16.

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

(3) Frankish J, Linden W: Is response adaptation a threat to the high/low reactor distinction in female college students? *Health Psychology.* 1991, *10:*224–227.

(4) Kamarck TW, Jennings JR, Debski TT, et al: Reliable measures of behaviorally-evoked cardiovascular reactivity from a PC-based test battery: Results from students and community samples. *Psychophysiology.* 1992, *29:*17–28.

(5) Turner JR: Inter-task consistency: An integrative re-evaluation. *Psychophysiology.* 1988, *25:*235–238.

(6) Fredriksen M, Dimberg U, Frisk-Holmberg M, Stroem G: Arterial blood pressure and general sympathetic activation in essential hypertension during stimulation. *Acta Medica Scandinavia.* 1985, *217:*309–317.

(7) Lamensdorf AM, Linden W: Family history of hypertension and cardiovascular changes during high and low affect provocation. *Psychophysiology.* 1992, *29:*558–565.

(8) Allen MT, Obrist PA, Sherwood A, Crowell MD: Evaluation of myocardial and peripheral vascular responses during reaction time, mental arithmetic, and cold pressor tasks. *Psychophysiology.* 1987, *24:*648–656.

(9) Mills PJ, Dimsdale JE: The promise of receptor studies in psychophysiologic research. *Psychosomatic Medicine.* 1988, *50:*555–566.

(10) Steptoe A: Theoretical bases for task selection in cardiovascular psychophysiology. In Steptoe A, Rueddel H, Neus H (eds), *Clinical and Methodological Issues in Cardiovascular Psychophysiology.* New York: Springer Verlag, 1985.

(11) DeLongis A, Folkman S, Lazarus R: The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology.* 1988, *54:*486–495.

(12) Suls J, Wan CK: The relationship between trait hostility and cardiovascular reactivity: A quantitative review and analysis. *Psychophysiology.* 1993, *30:*615–626.

(13) Fredriksen M, Matthews KA: Cardiovascular responses to behavioral stress and hypertension: A meta-analytic review. *Annals of Behavioral Medicine.* 1990, *12:*30–39.

(14) Prkachin KM, Mills DE, Zwaal C: Effects of acute stress and hostility components on blood cholesterol: The importance of social dimensions. (Abstract). *Annals of Behavioral Medicine.* 1993, *15*(Suppl.):S80.

(15) Light KC, Turner JR, Hinderliter AL, Sherwood A: Race and gender comparisons: II. Predictions of work blood pressure from laboratory baseline and cardiovascular reactivity measures. *Health Psychology.* 1993, *12:*366–375.

(16) Saab PG, Llabre MM, Hurwitz BE, et al: Myocardial and peripheral vascular responses to behavioral challenges and their stability in Black and White Americans. *Psychophysiology.* 1992, *29:*384–397.

(17) Manuck SB, Kasprowicz AL, Muldoon MF: Behaviorally-evoked cardiovascular reactivity and hypertension: Conceptual issues and potential associations. *Annals of Behavioral Medicine.* 1990, *12:*17–29.

(18) Barnett PA, Spence JD, Manuck SB, Jennings R: Psychological stress and the progression of carotid artery disease. *Journal of Hypertension.* 1997, *15:*49–55.

(19) Borghi C, Costa F, Boschi S, Mussi A, Ambrosini E: Predictors of stable hypertension in young borderline subjects: A five-year follow-up study. *Journal of Cardiovascular Pharmacology.* 1986, *8*(Suppl. 5):S138–S141.

(20) Perloff D, Sokolow M, Cowan RM, Juster RP: Prognostic value of ambulatory blood pressure measurements: Further analyses. *Journal of Hypertension.* 1989, *7:*S3–S10.

(21) Perloff D, Sokolow M, Cowan R: The prognostic value of ambulatory blood pressure monitoring in treated hypertensive patients. *Journal of Hypertension.* 1991, *9*(Suppl.):S33–S40.

(22) Parati G, Pomidossi G, Albini F, Malaspina D, Mancia G: Relationship of 24-hour blood pressure mean and variability to severity of target-organ damage in hypertension. *Journal of Hypertension.* 1987, *5:*93–98.

(23) White WB, Schulmann P, McCabe EJ, Dey HM: Average daily blood pressure, not office blood pressure determines cardiac function in patients with hypertension. *Journal of the American Medical Association.* 1989, *261:*873–877.

(24) Fredrikson M, Robson A, Ljungdell T: Ambulatory and laboratory blood pressure in individuals with negative and positive family history of hypertension. *Health Psychology.* 1991, *10:*371–377.

(25) Harshfield GA, James GD, Schlussel Y, et al: Do laboratory tests of blood pressure reactivity predict blood pressure changes during everyday life? *American Journal of Hypertension.* 1988, *1:*168–174.

(26) Ironson GH, Gellman MD, Spitzer SB, et al: Predicting home and work blood pressure measurements from resting baselines and laboratory reactivity in Black and White Americans. *Psychophysiology.* 1989, *26:*174–185.

(27) Langewitz W, Rueddel H, Schaechinger H, Schmieder R: Standardized stress testing in the cardiovascular laboratory: Has it any bearing on ambulatory blood pressure values? *Journal of Hypertension.* 1989, *7*(Suppl. 3):S41–S48.

(28) VanEgeren LF, Sparrow AW: Laboratory stress testing to assess real-life cardiovascular reactivity. *Psychosomatic Medicine.* 1989, *51:*1–9.

(29) Matthews KA, Owens JF, Allen MT, Stoney CM: Do cardiovascular responses to laboratory stress relate to ambulatory blood pressure levels? Yes, in some of the people, some of the time. *Psychosomatic Medicine.* 1992, *54:*686–697.

(30) Turjanmaa V, Tuomisto M, Fredrikson M, Kalli S, Uusitalo A: Blood pressure and heart variability and reactivity as related to daily activities in normotensive men measured with 24-hr intra-arterial recording. *Journal of Hypertension.* 1991, *9:*665–673.

(31) Ewart CK, Kolodner KB: Predicting ambulatory blood pressure during school: Effectiveness of social and nonsocial reactivity tasks in Black and White adolescents. *Psychophysiology.* 1993, *30:*30–38.

(32) Linden W, Con A: Laboratory reactivity models as predictors of ambulatory blood pressure and heart rate. *Journal of Psychosomatic Research.* 1994, *38:*217–228.

(33) Lai JY, Linden W: Gender, anger expression style, and opportunity for anger release determine cardiovascular reaction to and recovery from anger provocation. *Psychosomatic Medicine.* 1992, *54:*297–310.

(34) Earle TL, Linden W, Weinberg J: Differential effects of harassment on cardiovascular and salivary cortisol stress reactivity and recovery in women and men. *Journal of Psychosomatic Research.* 1999, *44:*1–17.

(35) Linden W, Earle TL, Gerin W, Christenfeld N: Physiological stress reactivity and recovery: Conceptual siblings prematurely separated? *Journal of Psychosomatic Research.* 1997, *42:*117–135.

(36) Linden W, Dadgar N, Earle TL: Cortisol responses to acute stress: A selective review (Abstract). *Psychosomatic Medicine.* 1994, *56:*153–154.

(37) Dienstbier RA: Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review.* 1989, *96:*84–100.

(38) Frankenhaeuser M: The psychophysiology of workload: Comparison between sexes. *Annals of Behavioral Medicine.* 1991, *13:*197–204.

(39) Kirschbaum C, Hellhammer DH: Salivary cortisol in psychobiological research: An overview. *Neuropsychobiology.* 1989, *22:*150–169.

(40) Abplanalp JM, Lingston L, Rose RM, Sandwisch D: Cortisol and growth hormone responses to psychological stress during the menstrual cycle. *Psychosomatic Medicine.* 1977, *39:*158–177.

(41) Bassett JR, Marshall PM, Spillane R: The physiological measurement of acute stress (public speaking) in bank employees. *International Journal of Psychophysiology.* 1987, *5:*265–273.

(42) Berger M, Bossert S, Krieg JC, et al: Interindividual differences in the susceptibility of the cortisol system: An important factor for the degree of hypercortisolism in stress situations? *Biological Psychiatry.* 1987, *22:*1327–1339.

(43) Hyppae MT, Aunola S, Lahtela K, Lahti R, Marniemi J: Psychoneuroendocrine responses to mental load in an achievement-oriented task. *Ergonomics.* 1983, *26:*1155–1162.

(44) Lane JD, Adcock RA, Williams Jr. RB, Kuhn CM: Caffeine effects on cardiovascular and neuroendocrine responses to acute psychosocial stress and their relationship to level of habitual caffeine consumption. *Psychosomatic Medicine.* 1990, *52:*320–336.

(45) Linden W, Long BC: Repression, hostility, and autonomic recovery from a laboratory stressor. *Journal of Clinical Hypertension.* 1987, *3:*567–578.

(46) Lovallo WR, Pincomb GA, Sung BH, et al: Caffeine may potentiate adrenocortical stress responses in hypertension-prone men. *Hypertension.* 1989, *14:*170–176.

(47) Meyerhoff JL, Oleshansky MA, Mougey EH: Psychologic stress increases plasma levels of prolactin, cortisol, and POMC-derived peptides in man. *Psychosomatic Medicine.* 1988, *50:*295–303.

(48) Miyabo S, Asato T, Mizushima N: Psychological correlates of stress-induced cortisol and growth hormone releases in neurotic patients. *Psychosomatic Medicine.* 1979, *41:*515–523.

(49) Pomerleau OF, Pomerleau CS: Cortisol response to a psychological stressor and/or nicotine. *Pharmacology, Biochemistry, and Behavior.* 1990, *36:*211–213.

(50) Sothman MS, Gustafson AB, Garthwaite TL, Horn TS, Hart BA: Cardiovascular fitness and selected adrenal hormone responses to cognitive stress. *Experimental Hypertension.* 1988, *7:*59–69.

(51) Wittersheim G, Brandenberger G, Follenius M: Mental task-induced strain and its after-effect assessed through variations in plasma cortisol levels. *Biological Psychology.* 1985, *21:*123–132.

(52) Cohen J: *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press, 1977.

(53) Dimsdale JE, Stern MJ, Dillon E: The stress interview as a tool for examining physiological reactivity. *Psychosomatic Medicine.* 1988, *50:*64–71.

(54) Uchino BN, Cacioppo JT, Malarkey W, Glaser R: Individual differences in cardiac sympathetic control predict endocrine and immune responses to acute psychological stress. *Journal of Personality and Social Psychology.* 1995, *69:*736–743.

(55) Asch S: Opinions and social pressures. *Scientific American.* 1955, *193:*31–35.

(56) Gerin W, Pieper C, Levy R, Pickering TG: Social support in social interaction: A moderator of cardiovascular reactivity. *Psychosomatic Medicine.* 1992, *54:*324–336.

(57) Linden W, Rutledge T, Earle TL: Gender differences and subtle protocol variations affect the replicability of anger provocations (Abstract). *Psychophysiology.* 1996, *33:*S10.