PHYSIOLOGICAL STRESS REACTIVITY AND RECOVERY: CONCEPTUAL SIBLINGS SEPARATED AT BIRTH?

W. LINDEN, T. L. EARLE, W. GERIN and N. CHRISTENFELD

(Received 30 November 1995; accepted 4 June 1996)

Abstract—This article discusses theoretical assumptions underlying physiological stress reactivity research. It examines early conceptualizations of activation and recovery and contrasts these with current practices in designing, analyzing, and reporting stress reactivity studies. Study protocols from four major journals covering the last 2 years of publication were examined for current practices. Of the 105 studies which tested physiological reactivity, 63% collected recovery data but only 23% reported the recovery data. We concluded that stress recovery issues are neglected and a renewed case is made for their conceptual and ecological importance. The case for studying recovery is further supported by a selective review of studies using recovery protocols that revealed positive findings not apparent in reactivity comparisons only. Finally, options for sound design of recovery protocols, statistical processing of resulting data, and interpretation of findings are presented. Copyright © 1997 Elsevier Science Inc.

Keywords: Physiological reactivity; Cardiovascular activity; Stress; Recovery; Emotions.

RATIONALE

This review article grows out of a perceived discrepancy between early and still frequently cited models of stress and disease and their operationalization in laboratory stress protocols with human subjects. Such discrepancies are held to weaken the heuristic value of the reactivity concept. The objective of this article is to review the conceptual grounding of stress reactivity research, to analyze current trends in the design of lab stress protocols, and to point out where theory and practice fail to come together. We present selective findings in an attempt to rekindle interest in studying recovery and discuss practical issues in the operational definition of recovery and related data processing and interpretation.

Because the terms reactivity and recovery are at the core of this study, brief definitions provided at the outset may be useful. Reactivity testing in the lab typically comprises an initial rest, or baseline period, followed by a period during which the subject is exposed to a stressor (like noise or an arithmetic challenge). Simultaneous physiological measures document the ensuing change which is then called reactivity. Recovery, in its simplest form can be defined as a poststress rest period that provides
information about the degree to which the elevation (i.e., reactivity) in the physiological and psychological parameters being measured persists after the stressor has ended.

**EARLY THEORIES**

In the study of stress–disease linkage, Selye’s contribution has been pivotal. As early as 1936, Selye [1] proposed to look at stress as a process involving three phases: activation; resistance; exhaustion. The body responds initially to challenge with physiological activation of a defense system, a resistance (or coping) phase follows during which stress is to be resolved, and if unsuccessful, the body may experience exhaustion. Activation that endures beyond the resistance stage is hypothesized to contribute to disease. In Selye’s model, the stress response is ubiquitous, involving all bodily defense systems. At about the same time, Freeman [2] posited that psychological recovery from experimental loads may be useful in estimating an individual’s ability to withstand conflict and related types of nervous strain in ordinary life situations. Such early suggestions that quick recovery from stress-induced arousal reflects effective coping are repeated in later theories of stress–disease linkage [cf. 3–5] and have never been challenged.

While Selye’s model implies a ubiquitous, “whole-system” response to stressors, the notion of response universality was later changed to distinguish at least two axes of physiological responding. The sympathetic–adrenal (SA) axis reflects activation due to motor and cognitive effort; when accompanied by adrenocortical hormone suppression this activation pattern has also been described as a “positive stress reaction” because it is short-lived and permits adaptive responding with maximal strength [6]. The hypothalamic–pituitary–adrenocortical axis (HPA) is thought to reflect affective distress and be the result of chronic, unresolved strain [3, 4, 6, 7], and it is also observed in anticipation of upcoming negative events [4]. SA activation is typically indexed by rises in epinephrine, norepinephrine, muscle tension, plasma free fatty acid levels, cardiac output, and blood pressure [8]. HPA axis activity is associated with increased circulating free fatty acids, suppression of immune function, and increased glucocorticoid production; HPA activation is inferred from the measurement of cortisol and its endocrine precursor, adrenocorticotropic hormone [4].

Drawing on animal and human studies, Frankenhaeuser [9] and Dienstbier [10] make a cogent case that the health consequences of HPA activation are notably more disease-relevant than are those of SA activation. In fact, Dienstbier’s conceptualization of a “physiological toughness” implies that transient arousal of a purely sympathetic–adrenal nature is unlikely to mark a pathway for disease.

Available research findings on aerobic fitness and acute physiological reactivity serve to support this point. There is consistent epidemiological evidence that physical fitness has disease-preventing and life-extending potential, yet achieving such fitness requires regular strenuous bodily activity marked by sympathoadrenal activation. Related studies of physiological reactivity to psychological challenge show that physically fit individuals do not respond any less than unfit individuals to psychological challenges; what distinguishes the fit from the unfit is the quicker cardiovascular recovery of the fit subjects [cf. 11, 12]. Dienstbier’s model handily explains why exercise can be a protective (inoculating) force that actually strengthens the
body's response capacity to disease and shows a pathway for the epidemiologically confirmed benefit of fitness on health.

Finally, there is recognition in the stress–disease literature that the study of physiological activation systems should include the cardiovascular, neuroendocrine, and immunological systems, and also investigate their interrelatedness [6, 13]. There is evidence that the cardiovascular system has myocardial and vascular activation components, the strength of which can vary from one individual to another, and these cardiovascular response patterns may also differentiate between types of psychological challenges [14]. The different defense systems should be recognized as interrelated but still allow for differential activation depending on the nature of the challenge (acute vs. chronic, cognitive vs. emotional, psychological vs. physical, to name a few). It is plausible that psychological challenges that trigger large responses in two or even all three defense systems provide more potent disease explanations than psychological conditions provoking change in only one system, and this may be especially true when that reactivity of only one system is accompanied by quick recovery. Frankenhaeuser [9] and Dienstbier [10] concluded that lasting changes in the neuroendocrine system are particularly health-threatening; and increasing evidence also suggests that this applies to the immune system [15]. Together, these theories prepared the conceptual ground for decades of stress research and they continue to sustain current research efforts. Whereas studies of stress and disease are recognized as difficult to interpret due to likely confounding of uncontrollable influences and the need to study disease outcomes over a long time, lab stress reactivity research has been considered a useful analog. It became a tool of choice because of its inherent advantage of control over moderating variables and relative ease of physiological measurement. By studying a resting or baseline state, experimentally intervening with a specific, quantifiable stressor, and by studying the ensuing physiological stress response, researchers intended to model stress responding in real-life and learn about the presumed stress–disease pathway.

To study the development of disease in humans would require single studies spanning many decades and are prohibitively costly and logistically challenging. This problem of human research was sidestepped by studying “windows” of activation; that is, time-limited arousal of physiological systems that are presumably involved in the disease process. The question, of course, remains as to how useful it is to limit one’s observations to what is observable through such narrow windows. One necessary condition for the ecological validity of the lab stress research approach is the demonstration that “manufactured” stress leads to an acute stress response and that it can lead to lasting physiological changes which actually mark the disease pathway in a causally plausible way. Even for the domain of cardiovascular reactivity (i.e., the most intensely studied stress response system) there is still only suggestive evidence for its predictive power in explaining disease [cf. 16] and even that suggestive evidence is seen by some as negligible [17].

How then can one study reactivity to a stressor in the lab while still believing that “stress” is being studied in a meaningful way? How likely is it that a stressor, even if it reliably triggers a physiological response, can be linked to a disease process when its effects disappear immediately upon withdrawal of the stressor itself? We posit that stressor protocols that allow for immediate recovery do not offer much explanatory power for studying disease pathways. There are two points to be made
here. Tasks that lead to very quick recovery may be contrasted with tasks that lead to slower recovery. This may provide some insight into the nature of the underlying processes that lead to sustained elevations in physiological activity, such as blood pressure. For example, Christenfeld and Gerin [18] have demonstrated that an interpersonal speech stressor produces poorer recovery than a mental arithmetic task. Thus they have hypothesized, interpersonal stress may lead to rumination and anger which may then keep the blood pressure elevated for a period after the stressor itself has ended; a mental arithmetic task, in contrast, may not. Anger triggered by the task could mediate sustained blood pressure elevations for example. The second reason for avoiding tasks that lead to uniformly swift recovery concerns the ability to examine individual differences in recovery following a stressor. Clearly, a task that produces no variability is not useful for such a purpose.

From this perspective, Haynes and his collaborators [19] undertook a review of what could be learned from examining physiological reactivity and recovery from stress. One clear conclusion from Haynes et al.'s [19] work is that lab challenges do not necessarily have uniform effects on reactivity and recovery, and that the failure to study recovery may prevent potentially exciting discoveries. Consequently, one would expect that recovery is routinely investigated in stress reactivity research. Next, we present results from a literature analysis that suggest otherwise.

**CURRENT PRACTICES IN STRESS PROTOCOL DESIGNS**

This article was motivated by the perceived neglect of recovery issues in stress reactivity research and we wanted to further substantiate this criticism with hard data from a review of current publication patterns. Two researchers (experienced postgraduate assistants) examined all issues of the 1993 and 1994 volumes of four prominent journals that frequently publish articles on stress reactivity research including physiological measures. These journals were *Psychosomatic Medicine*, *Journal of Psychosomatic Research*, *Psychophysiology*, and *Health Psychology*. In cases of doubt about whether a study should be considered a stress reactivity study, both researchers needed to agree with one another if a study was to be included. In cases of disagreement, the first author decided which studies to consider. The questions asked were: (a) How many articles described stress reactivity studies that included physiological measures? (b) What proportion of studies thus identified clearly stated that their protocol included a recovery phase during which physiological data were collected? (c) What proportion of studies actually analyzed and reported recovery data as they related to the overall experimental question? and (d) How was recovery analyzed?

For the 2 years in question there were 38 stress reactivity articles published in *Psychosomatic Medicine*, 22 in *Journal of Psychosomatic Research*, 34 in *Psychophysiology*, and 11 in *Health Psychology* for a total of 105. In 69 of the articles (63%) it is clearly stated that a recovery phase was part of the protocol; but in only 24 articles (23%) are recovery data actually reported. A summary of how studies defined and analyzed recovery data is provided in Table I.

As the data in Table I suggest there was extreme diversity in length of recovery phases and no consensus on best statistical treatment. Noticeably, the studies with endocrinologic and immune data used much longer recovery periods than those with only cardiovascular data. Most likely, these researchers were already cognizant
Table I.—Protocol characteristics of reactivity studies with recovery periods

<table>
<thead>
<tr>
<th>Type of endpoints</th>
<th>No. of studies</th>
<th>Length of recovery (min)</th>
<th>Frequency of chosen statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>Sympathetic</td>
<td>13</td>
<td>9.4</td>
<td>2–30</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indices</td>
<td>10</td>
<td>57.5</td>
<td>5–240</td>
</tr>
<tr>
<td>Endocrine indices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunological</td>
<td>1</td>
<td>20</td>
<td>N/A</td>
</tr>
<tr>
<td>indices</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

None = recovery data not analyzed, only means reported; RM = repeated measures ANOVA; ANC = analysis of covariance; CS = (simple) change scores; RCS = residualized change scores; other = paired t test used.

of the relatively slow response of endocrine indices thus requiring long observation periods. Note also that almost one half of researchers reported recovery raw means but did not analyze recovery data. The primary reason given for why recovery data were collected but not reported or analyzed was that recovery was generally swift and that analyses did not appear promising.

When two thirds of researchers fail to deal with recovery the question arises as to why recovery is so often ignored. There are several possible reasons. For one, the sort of tasks typically used in the lab may allow for such rapid recovery that the researcher cannot detect individual differences in recovery or different rates of recovery for varying tasks. Experience has born this out: between us, we have published more than 20 studies on physiological stress reactivity; the tasks we used included role-play of assertion-requiring social situations, counting forward in steps of one, Stroop tests, math tests with various kinds of noise interferences or anger provocation, anger recall tasks, cold pressor, discussion of controversial topics in small groups, and more. We learned that the great majority of subjects (even those at risk because of elevated baselines values, or because of positive family histories) displayed return-to-baseline levels within 1–2 minutes upon termination of the stressor. Swift recovery was typical for most lab tasks with the notable exception of tasks provoking anger; in these instances, even 10-minute recovery periods were not sufficient to permit group averages of blood pressure to return to baseline levels [20, 21].

The typically swift recovery may well be an artifact of the laboratory: the experimenter will select tasks over which s/he can exert complete temporal control; that is, when the precise instant at which the stressor begins and ends can be manipulated. In the real world, of course, stressors tend not to have distinct beginnings and endings. Also, subjects are less likely to engage in contrived lab tasks as much as they do in real life, and we think it is crucial that researchers understand the emotional/psychological experience of their subjects in the lab [22]. Finally, as we discuss next in more detail, recovery may be ignored by many researchers because it poses greater measurement problems than do reactivity protocols.

DISCOVERIES DERIVED FROM RECOVERY PROTOCOLS

Animal studies

Much of our knowledge of the physiology and endocrinology of hypertension, other cardiovascular diseases, and cancer has come from animal research. Several
animal models suggest homology with the human disease process [23]. A selective review of animal studies revealed that researchers have often taken full advantage of the uniqueness of the animal model by carrying out extended protocols involving initially healthy animals who are followed until the development of actual disease. Hypertension is one particularly well studied area and can serve as an exemplar of the animal model approach. Several types of animal models of stress-induced hypertension exist in the literature [24]: (a) Aversive conditioning models were used to create hypertension over 6–7 months by shaping the animals’ behavior [24]. (b) Psychosocial stress has been operationalized as intraspecies dominance hierarchies; mice raised in all-male colonies, which develop dominant behavioral patterns, developed high blood pressure over 4 months [24]. (c) The breeding of spontaneously hypertensive rats (SHR) [25] allowed research showing greater cardiovascular reactivity to even mild environmental stressors, such as light and noise in SHR.

While not attempting to comprehensively review this literature, one clearly shared characteristic of animal studies is that stress is operationalized as a chronic condition, animals are studied over many months, and disease processes from presumed triggers to actual disease outcome are actually studied.

Review of animal studies also revealed that many researchers fruitfully exploited the greater ease with which multiple physiological systems can be monitored in animals and this appears especially true for measures of neuroendocrine activation and recovery [26–29]. Anisman et al. [29] reviewed explanations for the observed pattern of attenuated endocrine recovery and concluded that, as long as the stressor is ongoing, increased hormone levels are necessary for coping; however, continued elevation beyond termination of the stressor is seen as an “anticipatory readiness” for future stressors, a response conditioned by the stressor itself and related cues. If the stressor is unpredictable, endocrinological/physiological adaptation to the stressor is slower to occur. In such paradigms, wherein the animals have no behavioral means of controlling, predicting, or terminating the stressor, generally several endocrine indices show sustained change. Depending on the length of the paradigm (if using more than one 24-hour session), uncontrollable stress typically results in depletion of norepinephrine, and increased HPA indices (ACTH, corticosterone, corticotrophin-releasing hormone, cortisol). When learned helplessness is created via social stressors (i.e., separation of young primates from their mothers), HPA indices take several days to begin to return to normal parameters (i.e., adaptation to the chronic stress is very slow).

On the whole, animal research has taken a direct, process approach to studying stress and disease. Because of the longitudinal approach, the typical definition of stress as a chronic condition, and the ongoing observation, animal researchers did not need to concern themselves with precise definitions of baseline, reactivity, and recovery; also, the ecological validity of their stressors is more obvious than it is in human research.

Studies with human subjects

Experimental studies of stress and recovery with human subjects can be subdivided into those where: (a) studying recovery strengthened the results already observed during the baseline-task phase; (b) interesting findings were apparent only
during recovery but not during the baseline-task phase; and (c) researchers targeted recovery processes \textit{a priori} as the objects of study. A good example of the first kind is the study by Fleming et al. [30] pooling naturalistic and contrived lab stress environments. Fleming and Baum used the embedded figures task to study lab reactivity in individuals who were predicted to react differently to stress. One group was defined as high chronic stress because they lived in crowded environments, the other was a low chronic stress control sample living in uncrowded neighborhoods. The researchers found the predicted magnifying effect of the underlying chronic stress situation on blood pressure and heart rate reactivity to the lab task. In addition, they noticed that although this type of task typically allows prompt return of cardiovascular indices to baseline (as seen in their control group) there was a systematically slower recovery in the chronic stress group. In a similar study, male factory workers were unwinding more quickly from acute stress after their vacations than before [31].

Arguably the best example of studies showing interesting results only during recovery are those on the arousal-reducing benefits of aerobic fitness; investigation of the recovery process has turned out to be pivotal in this case. As for cardiovascular benefits of aerobic fitness, most studies fail to show a benefit of fitness on the predicted attenuation of the cardiovascular stress response when only baseline to task changes are measured. But, it has consistently been found that recovery from various lab challenges is quicker for the aerobically fit [cf. 11, 12]. These data suggest a plausible pathway for the health-protective effects of aerobic fitness.

There are many other examples where the study of recovery was the most revealing part of a lab stress protocol. One type of task where delayed recovery was particularly strong appears to be any challenge that involves anger provocations [20, 21, 32, 33]. Studies spanning a three-decade timeframe consistently revealed large physiological responses to anger provocation that exceeded responses typically observed with other, non-anger-related lab challenges. Hokanson and collaborators [32] discovered that men’s blood pressure increased in response to anger provocations and that they recovered more quickly when they had a chance to express their anger relative to “being stuck with it.” In contrast, women did not show quicker blood pressure recovery when expressing anger relative to not expressing it; and, similarly, it was not physiologically beneficial for either women or men to express anger toward a superior. Whereas Hokanson focused on situational manipulations and their acute effects, successive studies involving his paradigm studied the interaction of individual difference, trait-type factors like anger expression preferences [20, 33] and acute anger provocations. In essence, these researchers observed accelerated recovery in those men where situation and personality matched. For example, an anger-out subject given a chance to express his frustration recovered more quickly than a subject preferring an anger-out style who was experimentally constrained from expressing his feelings. Interestingly, while these findings suggest that recovery may be affected by an interaction of both situational factors (opportunity to express anger) and anger expression styles (seen as stable traits), this interaction was more often seen in men.

Earle et al. [21] investigated anger responding in men and women, used cardiovascular and cortisol measures, and included groups performing a challenging math task with and without harassment. To the best of our knowledge, this was the first
time that acute anger provocation was tracked in terms of SA and HPA activation. The addition of harassment doubled the response magnitudes of math-only challenges. We also replicated the previously observed slow cardiovascular recovery from anger, and the observation of a large cortisol response and slow cortisol recovery in harassed men but not in women.

In summary, anger provocation tasks reflect types of stressors from which subjects do not routinely show full return to baseline recovery within 1–2 minutes; even at 10 minutes post-task only about half the response magnitude is typically recovered. Furthermore, consistent gender differences were noted: harassment affected men’s physiology much more than women’s. Wherever cardiovascular recovery was slow there also was a magnified acute cortisol response and slow recovery suggesting anger-induced HPA axis activity. Findings from these studies are entirely consistent with a massive body of epidemiological work suggesting that chronic anger and hostility may contribute to CV disease [34–36].

The investigation of affect in the broader sense has also provided interesting findings. Using older adults, Vitaliano et al. [37] exposed their sample to a cognitive and an emotion-provoking task and studied their cardiovascular responses. Subjects recovered more slowly from the emotion task than from the cognitive tasks. Vitaliano et al.’s work also deserves special mention because they studied not only the effect of task choice on reactivity and recovery, but they also identified individual-difference factors that are associated with differential recovery. Regression analyses indicated that slow recovery was associated with elevated anxiety scores, avoidant coping styles, and low scores on anger held-in. Furthermore, these researchers noted that men recovered more slowly than women, and marked obesity was also associated with attenuated recovery. Subjects on antihypertensive medication recovered more quickly.

Another important contribution to the recovery literature was made by Borghi and collaborators who conducted a 5-year follow-up of young borderline hypertensives [38]. These researchers found that diastolic blood pressure (BP) recovery from a cognitive challenge was a more useful predictor of long-term BP changes than was reactivity itself. Similarly, slow recovery following physical activity had predictive validity for the development of hypertension [39].

A clever expansion of the baseline, stress, recovery lab protocol was used by King et al. [40], and Warren and Pieper [41] who conducted ambulatory blood pressure monitoring studies involving work and home environments. In both studies, many of the contingencies of home and work are known (e.g., home is considered equivalent to the recovery phase in a lab experiment) and the subject’s experience is to some degree open to individual control. No doubt arises about the ecological validity of the events studied in ambulatory protocols. In King et al.’s [40] study, the subjects were caregivers to a chronically ill person in their home and it was found that return from work actually increased stress as determined via change in cardiovascular activity levels relative to their activity levels while at work. These findings were further strengthened by the observation that returning home for noncaregivers led to reductions in blood pressure levels. Similarly, the Warren and Pieper study [41] involved a typical work day and a subsequent day off work; these researchers identified an absence of the otherwise typical “unwinding” in subjects who were at greater cardiovascular risk. Frankenhaeuser described similar studies
comparing men and women's work and home situations [9]. Typically, men showed decreasing stress levels (i.e., indexed by catecholamine determinations) when returning home, whereas women, and especially those with children at home, showed increasing stress levels when returning home. These latter studies represent a good blend of ecological validity and methodological control and their results consistently point to the importance of studying recovery with a lengthy observation period.

INDIVIDUAL DIFFERENCES IN RECOVERY

As many of the above studies suggest, there are many individual differences in the speed of physiological recovery. While we have not provided an exhaustive review, it appears age, medication, and anxiety may negatively affect recovery. Gender differences suggest a complex pattern such that men recover more slowly than women in the laboratory; however, this gender difference may disappear or even reverse when the stressor is real-life and individually relevant. It therefore makes sense not only to identify individual-difference factors that account for differences in acute response to a stressor but to test as well whether the same individual-difference factor also accounts for differential rates of recovery. It is at least plausible that an individual-difference factor that increases reactivity magnitude also slows down recovery, and this type of individual difference may be of particular importance for the study of etiological processes. Similarly, there may be numerous individual difference factors that act like aerobic fitness in that they do not have a main effect on reactivity but may affect recovery.

RECOVERY PROTOCOLS AND DATA ANALYSIS ISSUES

The following sections summarize some of the issues to consider when designing experiments with physiological reactivity and recovery protocols. Although we recognize that experimental design and data analysis are two sides of the same coin, these issues shall be considered separately, for the sake of exposition. Note that the following is not intended to be an exhaustive review of how to conduct reactivity studies (we recommend Steptoe [22] or Turner [42] for this purpose).

On the surface, the development of recovery protocols seems relatively simple, especially within the laboratory setting: one simply adds a poststress resting period to a standard reactivity protocol and continues to monitor physiology, behavior, etc. In developing such protocols, however, several issues must be considered. Some of these concern both the reactivity and recovery aspects of the protocol; others are relevant to the recovery phase only.

Test-theoretical issues

Just as reactivity studies are psychophysiological tests, so are recovery studies; basic criteria for reliability and validity need to be met.

(a) Recovery must be reliably measurable if it is to have any predictive power. There are two meanings of reliability in this context. (i) It may reflect an individual difference, which means that test–retest correlations using the same sample should be strong. This is necessary if the recovery measure is to be regarded as an individual difference risk factor (such as race or family history of a disease). (ii) The second
meaning of reliability in this context is that differences between conditions (i.e., main effects) can emerge, regardless of the rank order of individuals within each condition. For example, if a group of subjects participating in an anger-recall task is either given the opportunity to continue to ruminate about the cause of their anger following the task, and another group of subjects is distracted in some manner following the task, a reliable measure of recovery might be expected to differentiate these two groups. In this case, the findings would be relevant to how situations could be altered to reduce stress (and possibly its negative health consequences) even if the measure could not be used to identify individuals who are at greater or lesser risk for the development of cardiovascular-related disease.

(b) *It has psychophysiological significance.* Psychophysiological significance can be accomplished by providing a measure of the operation of some physiological system or process (e.g., metabolism rate of activating hormones or arterial compliance). In addition, this measure might provide insight into cognitive/emotional mechanisms (e.g., effects of anger). Alternatively, the recovery measure might show psychophysiological significance by demonstrating power to predict later pathology. These probably often overlap, but not necessarily.

(c) *It should generalize beyond the laboratory.* If the recovery measure is useful only in the laboratory (following tasks such as the cold pressor or mental arithmetic), and does not generalize to recovery from (or some other correlate of) real-world stressors, then it is unlikely to play a major pathogenic role in the development of disease, or even to describe the normal operation of the system. People may show reliable individual differences in physiological recovery from cold pressor episodes, for example, that reflect nothing more than how they feel about cold water. And it may not generalize to the rate of recovery from stress such as public speaking, anger at a spouse, a stressful work day, or even jogging in place.

(d) *The recovery index should provide an increase in information over the measures now used.* There are two ways recovery could serve this purpose. One is to provide additional information about the same underlying physiological and/or cognitive processes that are already assessed in studies of reactivity; the other is to provide some measure of processes independent of those tapped via the reactivity measure. If the index of recovery is simply another way of tapping the activation of the system during the stressor, it may simply be duplicating the information provided by the reactivity measure. This is of course an empirical question and we were able to provide numerous examples which underscore the unique value of studying recovery. To some extent, reactivity and recovery must be related, because if the stressor has absolutely no effect on the parameter being measured, for example, diastolic blood pressure, then there is nothing to recover from. Even if the measures are related, however, recovery information will still be useful if it turns out to be more reliable and/or generalizable than the reactivity measure alone.

**Practical design issues**

*What is the nature of the parameter being measured?* Some measures, such as heart rate, blood pressure, and blood serum levels of catecholamines are affected almost immediately by mental and physical stressors [42]; other responses, such as cortisol, may take 20–30 minutes to peak [43]. Thus, while blood pressure can change measurably in less than 60 seconds, cortisol must be assessed over a longer
Review

What is the sampling interval for the parameter under investigation? Sampling interval issues bear on the reliability and sensitivity of the reactivity/recovery measures [44]. By taking advantage of a simple statistical principle, given by the Spearman–Brown formula for stepped-up reliability, increasing the number of measures increases reliability. For example, in a typical blood pressure reactivity procedure, measures are taken in 60–120 second intervals and the researcher often has little choice in the matter because of constraints imposed by the measurement tool. In this case, patterns within the sampling period cannot be investigated. Alternatively, if sampling is very frequent (e.g., taking beat-to-beat measurements of heart rate or stroke volume), such a pattern becomes accessible for examination. This is particularly important for poststress recovery periods where the presumption is one of rapid change, typically a decrease. We often found cardiovascular recovery to be so swift that a 2-minute measurement interval misses the entire downslope.

What is an appropriate definition of baseline? Surprisingly few studies have addressed issues in the assessment of baseline [19, 45, 46]. This is not a trivial question for recovery researchers. Haynes et al. [19] define recovery as return to baseline. This would make sense if baseline definitions were consensual and possessed a clinical meaning, but neither of these conditions holds. Also, during recovery, physiological activity indices may fall below levels used as baselines. While it is unknown at this time whether such post-task compensation is of clinical significance it cannot be studied without careful recovery observation. Note that there is great variability between laboratories in the duration of the initial rest period, and in how to allow for adaptation [45]. At least for cardiovascular data, it has been established that completion of questionnaires, magazine reading, and quiet rest equally allow for a similar degree of adaptation. Generally a 15–20 minute adaptation period appears sufficient to obtain a relatively stable plateau in blood pressure that permits averaging for a reliable baseline definition [47].

Obrist has suggested, for example, that subjects be brought into the laboratory on one day, and a baseline assessed; and then be brought back on a different day, to respond to the stressor [48]. Some researchers claim that “baseline,” as taken in reactivity/recovery studies, does not represent a true resting level, but only a relative value, which at best may be used for comparison with the stress level, as in the computation of a stress-baseline change score. In this vein, Jennings et al. [49] recently published a study in which they suggested the use of a “vanilla baseline,” that is, one during which the subject is occupied with a low-demand cognitive task. The rationale for this is that the mental processes of the subjects will at least be somewhat controlled (across subjects), in contrast to the usual procedure in which subjects’ thoughts are allowed to wander as they might. A variation of this procedure allows the subjects to look at magazines (presumably unexciting ones) during the initial rest period.

What should the subject be doing during the recovery period? Issues in the determination of an initial baseline also are relevant for the recovery phase. In fact, “recovery,” if followed by a second stressor, also may play the role of interstress baseline.
Interestingly, in studying recovery, researchers are focusing on the same phenomenon they tend to label and dismiss as “adaptation” in the initial baseline. The only difference between these is that the activities preceding adaptation in the initial baseline are left uncontrolled, compared to those preceding recovery, which of course is response to the stressful task [19]. What the subject is actually doing during the poststress recovery period depends partly on the purpose of the investigation. For example, if one is interested in the effects of anger on poststress recovery, the degree to which the subject tends to continue ruminating about the preceding stressor may influence the recovery curve, and interfere with the examination of “pure” residual effects of anger. In this case, it would make sense to distract the subject during the recovery period.

DATA REDUCTION/ANALYSIS

This section deals with issues that arise in data reduction and statistical analysis. As will be shown, analysis of recovery data is beset with all the problems of analyzing reactivity plus some unique ones. A brief summary of the options for analyzing change is presented first.

Change scores

Change scores are appealing because they are easy to calculate and understand. Psychometricians have long been aware of the limitations of change scores—chiefly, the problem of reliability of the change score, which depends on the reliability of each of the component levels (i.e., baseline, stress), as well as on the correlation between them [50]. In reactivity studies, the usual change score is simply the task level minus the baseline. Averaging across multiple measures of a baseline period and a task period increases reliability. Another solution has been suggested by Kamarck et al. [51], who suggest the computation of a mean change score, based on the change scores in response to several, related stressors. If the researcher can show that the baseline value does not correlate with the subsequent change, if that is equally true for all groups in a multigroup design, and if these groups are not different from one another at baseline, then change score analysis is useful and an elegantly simple solution. Unfortunately, these conditions are very unlikely in recovery designs.

Repeated measures

Some researchers use a repeated-measures design of analysis, in which the baseline and stressor levels are treated as two levels of a single factor (i.e., “phase” or “time”). However, this procedure does not get around the problem noted with change scores; in fact, the effects of the between-subjects factors will be identical for these two designs.

Analysis of covariance (ANCOVA) and residualized change scores

Sometimes baseline levels are used as covariate measures. Manuck et al. [52] suggested that such a procedure follows from the view that the residual variability, after adjusting for baseline, is due to the task, and not to individual differences in baseline levels. One decision rule that is often employed is based on the association
between the baseline measure and the change score. If the correlation is substantial (i.e., $r > 0.6$) then baseline measures may be included as covariates in the analysis. Note, however, that the researcher must demonstrate parallel regression slopes for all groups in a multigroup design before ANCOVA is justified. For blood pressure, Manuck et al. [52] reported that a lack of association for the baseline and the subsequent reactivity is commonly found in studies involving normotensive subjects. This makes sense, because, by definition, such a sample will have a restricted range. If the sample comprises both normotensive and hypertensive subjects, the correlation between baseline and change should be greater. In this case, it is important to note that individuals with higher baseline levels will show greater reactivity than those with lower baselines. This issue is reflected in the controversial issue concerning whether normotensive–hypertensive differences in reactivity should be measured in absolute terms, or in percentages of baseline values. Several studies indicate that hypertensives evidence greater reactivity than normotensives [53]; however, this effect disappears in many studies when the percentage, rather than the absolute level, is considered [54]. Once again, the “correct” measure depends on what pathogenic model of reactivity is invoked.

An alternative to ANCOVA gaining in popularity is the use of residualized change scores where the baseline adjustments are not based on the regression slope of the whole group but where the impact of baseline on reactivity is adjusted for each individual. While considerably more laborious than ANCOVA, this appears to be the least problematic procedure.

**Recovery outcome measures**

While all measurement of change remains somewhat controversial, the question of an appropriate measure of recovery is a more difficult one. It is desirable for a recovery measure to be largely independent from the reactivity measure, to insure that we are not just getting a different measure of the same dimension.

To validate recovery from stress as an important risk factor for the development of disease, an additional set of criteria must be employed. Namely, the measures should pass the same tests as have been applied to reactivity: the measure should be sensitive to differences between groups (e.g., with normotensive and hypertensive status); should reflect influences of family history of a disease (we have provided some evidence that this occurs [55]; and should predict the future development of disease. Because few of these criteria—associations with other risk factors and disease endpoints—have been tested using recovery models, we are not yet in a position to judge the suitability of individual recovery measures on the basis of how useful they will be for this purpose. The following section describes some of the measures that might be used for assessing recovery, and their advantages and disadvantages. However, it is important to note that the suitability of some measures over others is not made with reference to clinical outcomes.

**Time to recovery**

This measure seems to appeal to the intuition of how recovery ought to be measured but there are several problems. The main concern is that the parameter may not return, within the allotted period. At some point, a “recovery” period, which theoretically provides a condition under which resting levels will be reacquired, can
become a stressor in and of itself, due to the boredom and restlessness which may occur. We have, on occasion, observed that: (a) blood pressure did not return to prestress levels within 20 minutes; and (b) by the end of a 20-minute recovery period, blood pressure levels actually rose again [19]. Thus, recovery, by this definition, may not occur, and one is left having to assign some arbitrary ceiling value, a measurement practice which should usually be avoided. A second problem with the “time to recovery” measure is that it ignores the slope of the recovery curve. For example, two subjects may both evidence heart rate of muscle tension levels that return to baseline in, say, 10 minutes; however, one of the individuals may immediately decline almost, but not quite, to the point regarded as recovered, and remain elevated barely above this point for the remainder of the 10 minutes. The other individual, however, may only begin to decline during the final few minutes, to be labeled as “recovered” by 10 minutes. Both of these subjects show the same recovery, by the “time to recovery” definition; however, common sense suggests that they are not. One final problem: a measure of time until a single reading falls within some specified range is likely to be highly unreliable with measures such as blood pressure and heart rate, which are highly variable.

**Area under the curve**

The measure of area under the curve is an improvement on the one-dimensional “time to recovery” measure, as it does control for the steepness of the decline in the level of the physiological parameter (i.e., the second objection noted previously). It is noteworthy that people who fail to recover within some specified period will not throw this measure off. Two (resolvable) problems with examining area under the curve are: (a) it requires a frequent sampling interval; and (b) it is likely to be influenced by the initial level of reactivity. The second problem could be tackled by simply looking at the recovery phase as part of the same baseline task protocol and by calculating a single area under the curve index that covers the entire time span from the baseline to the last measurement point.

**Poststress level at arbitrary intervals**

This is probably the most widely used measure in studies that have assessed recovery. Using this method, the researcher takes a measurement at one or more arbitrary poststress intervals (e.g., at 5 and then at 10 minutes following stressor offset). This technique holds the advantage that intermittent measurement techniques can be employed (e.g., a single blood pressure measurement can be taken); and does not suffer the problems connected to the use of change scores. Given how little is known about post-stress recovery patterns, and their clinical implications, this method relies on the experimenter’s intuition concerning an appropriate sampling interval. For example, 5 minutes after stressor offset seems like a reasonable point at which to collect recovery data, but in fact the crucial period, if one exists, might occur even sooner or later. Finally, in a related vein, this method obscures the overall pattern of recovery. Ten minutes poststressor, two subjects may find themselves at the same physiological level, but the means by which they both got there remain unknown.

Once again, the issue of sampling intervals is relevant here. We suspect that one reason why poststress recovery has remained understudied is quite pragmatic: for
measures such as blood pressure, most researchers have been limited to an apparatus that allows sampling intervals of 1 minute, at a minimum. Such an interval does not allow one to produce much in the way of a descriptive recovery curve. An alternative, for blood pressure measurements, is to use either intra-arterial methods (which are invasive, and therefore add several complications to the protocol) or noninvasive methods, such as the Finapres, which has proven useful in these designs [44].

**Poststress (at arbitrary intervals) minus baseline change scores**

This method holds the same basic advantages and disadvantages as the use of the poststress level at arbitrary intervals, with one additional advantage. This measure appears intuitively useful, because it provides a measure of recovery relative to the prestress value. For example, let us suppose that a particular subject has a baseline diastolic blood pressure of 60 mmHg; during stress, the diastolic pressure rises to a peak of 80 mmHg; and, at 5 minutes after the stressor offset, the subject's diastolic pressure is 65 mmHg. The baseline to poststress (recovery) change score here is 5 mmHg over baseline.

Of course, the advantages and disadvantages of using the change score described at the beginning of this section also hold true here. Thus, this score is subject to reliability problems. However, in terms of comparison between subjects, the change value may be more meaningful than the absolute level.

**Poststress (at arbitrary intervals) minus stress level change scores**

These differ from the previous measures in that the change score provides information relative not to the prestress level, but to the stress level. Thus, the change score is interpretable as the degree to which the parameter has declined (or increased) since the termination of the stressor. Using the previous diastolic blood pressure example, a subject who evidences a 60-mmHg prestress level, an 80-mmHg stress level, and (at 5 minutes) a 65-mmHg poststress level, would therefore show a 15-mmHg decline in diastolic pressure.

It is worth noting that both change from baseline and change from stress levels are likely to be highly influenced by the degree of reactivity. To control for such differences, one might consider using the percent recovery, rather than absolute levels, as a measure. This is analogous to the debate between use of change scores (measured in absolute units) versus percent change from baseline of blood pressure response to a stressor, in comparisons of reactivity between normotensive and hypertensive subjects. Some investigators have found, for example, significant differences between mean reactivity change scores for these groups but the difference disappears when a percent change measure is used [54].

**Curve-fitting estimates**

Another approach to capturing the nature of the recovery process is to use more sophisticated curve-fitting techniques that describe the recovery process with a mathematical equation. These procedures have a number of advantages. Curve-fitting, using a least-squares algorithm, provides an estimate of the parameters based on all of the relevant data points, rather than a single point or an arbitrary subset, and should therefore show greater reliability: outliers will do less damage. Another
advantage is that they can provide estimates of several parameters simultaneously. If, for example, more than one physiological process is involved in poststress recovery of prestress baseline, it is possible that these could all be represented with a single curve for the sake of parsimony. An example follows: If the majority of heart rate recovery occurs within 1 or 2 minutes of the stressor offset, and the heart rate then remains elevated for a long period at a level that is slightly above the previous baseline, one could estimate both the slope of the initial recovery and the asymptotic long-term recovery level. Third, curve-fitting approaches can be made independent of the level of reactivity (e.g., the slope of the recovery is not necessarily a function of the amount of recovery that occurs). In addition, such a measure is unaffected even if the subject never returns to a prestress level.

RECOMMENDATIONS FOR FUTURE RESEARCH

We hope that this article kindles (or rekindles) interest in recovery. The study of recovery is a logical consequence of theories on stress–disease pathways and has revealed some important findings. For example, many researchers [9, 10] have stressed the importance of HPA axis activity in understanding stress–disease links and HPA activity cannot be meaningfully studied without lengthy recovery protocols.

Generally, physiological recovery appears slowest when tasks provoke anger, and recovery was often slow when studies were conducted under naturalistic stress conditions; this may have important disease implications and suggests fruitful avenues for future research. It is further noteworthy, that the physiological impact of interpersonal stressors clearly mirrors the results from other stress studies where it has been shown that humans habituate quickly in their emotional tone to many stressors but they do not habituate well at all to emotional distress such as marital discord [56]. In contrast, the majority of typically used lab tasks allows for immediate recovery and are not likely useful tasks in studying stress–disease linkage unless it can be shown that recovery is swift in one physiological system (let us say cardiovascular), but lasting activation in another system (e.g., the immune system) is shown at the same time [57].

A change in publication practices can help to maximize the usefulness of studying recovery. We urge those who conduct reactivity studies to take poststress measurements, and to report even null results. This will assist the research community in developing useful (and avoiding problematic) approaches to develop this paradigm. For the sake of study comparability, we urge for a consensus (via editorial decisions, a consensus conference, etc.) on how to ideally define baselines and recovery, including consensual decisions on statistical analyses. An example of this is the editorial policy of the journal Psychophysiology on the analysis of repeated measures [58]. At this time, we suggest to sample frequently, and wherever possible, continuously, for maximal test reliability. We also believe that researchers should strive to study SA and HPA axis activity simultaneously, and that studies testing for HPA activity should have post-task observation periods of at least 30 minutes, preferably 60 minutes [59].

Acknowledgments—Preparation of this article was supported by operating grants from the B.C. and Yukon Heart & Stroke Foundation and the Social Science and Humanities Research Council of Canada
awarded to the first author, by a career training award from SSHRC awarded to the second author, and a New Investigator award from NIH to William Gerin. The thoughtful criticisms of Joseph W. Lenz and Thomas Rutledge on an earlier version of the manuscript are greatly appreciated.

REFERENCES


