

ASSESSMENT OF STRESS

# TOWARD THE STANDARDIZATION OF LIFE STRESS ASSESSMENT: DEFINITIONAL DISCREPANCIES AND INCONSISTENCIES IN METHODS

JOHN R. McQUAID, MS AND SCOTT M. MONROE, PhD  
*University of Oregon, USA*

AND

JOHN R. ROBERTS, MS, SHERI L. JOHNSON, MS, GREGORY L. GARAMONI, PhD, DAVID J. KUPFER, MD  
AND ELLEN FRANK, PhD  
*University of Pittsburgh, USA*

## SUMMARY

There has been considerable controversy about methods for assessing life stress. However, self-report checklists and interview-based measures (the predominant approaches used in current research) differ in several respects, ranging from basic definitions through theoretical assumptions. Most research comparing these two approaches has focused on global comparisons in predicting disorder, which fail to take into account more specific information on how the methods vary. The present article outlines three stages of assessment for life stress: definition, operationalization, and quantification. Detailed examination of these stages with a sample of depressed patients helps to demonstrate in an explicit manner how self-report checklists and interview-based methods differ at successive stages of the measurement process. Data are presented that indicate large endpoint discrepancies attributable to specific differences in the definitional and operational procedures used in the two assessment approaches. The nature of the discrepancies found is discussed, along with the implications for assessing life stress and testing its implications for health and well-being.

**KEY WORDS**—Stress, life events, chronic stressors, assessment.

A small number of studies have directly compared self-report checklists and interview-based procedures in the assessment of life stress. The conclusions from such efforts have been as varied as the methods employed. Zimmerman *et al.*<sup>1</sup> examined the reporting of depressed patients and found that 38.8 per cent of the items elicited in an interview procedure had not been reported on a checklist, and 41.3 per cent of the events reported on the checklist were not assessed in the interview.

In a similar type of study,<sup>2</sup> Oei and Zwart found that there was a significantly different number of events reported on a checklist as compared to an interview. Specifically, depressives reported on the questionnaires more work problems, education events, personal illnesses, and marital problems; on the interview they reported more deaths. Katschnig<sup>3</sup> also compared life events reported by depressed patients using both checklist and interview-based procedures. The findings showed a relationship between both types of measurement and symptomatology. Remarkably, though, a closer analysis found that the presence or absence of an interview event was uncorrelated with the presence or absence of a

Address for correspondence: John R. McQuaid or Scott M. Monroe, Department of Psychology, Straub Hall, University of Oregon, Eugene, Oregon 97403-1227, USA. Phone: (503) 346-4921.

Table 1—Three stages of assessment in the definition and measurement of life stress

	Stage 1	Stage 2	Stage 3
<u>Descriptor</u>	<u>Definition</u>	<u>Operationalization</u>	<u>Quantification</u>
Activity	Delineating domain of experiences to sample	Translating experiences into different categories	Combining categories into summary measures
Major guiding principle	Theory	Standardization	Theory

checklist event ( $\kappa=0.04$ ). Katschnig concluded that there were two possible explanations for the results. Either the two instruments were measuring different but relevant phenomena or, due to low reliability, the SRE was producing spurious results.

These studies suggest that there are significant differences between the information provided by self-report and interview procedures for assessing life stress. However, self-report checklists and interview-based procedures differ along several lines which may have distinct implications for operationalizing life stress and for studying its implications for disorder. General comparisons between the two approaches are likely to obscure more specific differences in the definition and measurement of life stress. The present study examines how the self-report and interview-based measures of life stress differ on several underlying dimensions.

It is helpful to distinguish three stages in the assessment process that place such research in a more interpretable context (see Table 1). Stage 1 involves decisions regarding the basic nature of the information initially elicited: the definitional stage. What are the basic phenomena of interest? Is the spectrum of life's daily activities included, or are distinctions made in terms of magnitude (e.g. major versus minor events), dimensionality (e.g. desirable versus undesirable events) or chronicity (e.g. acute versus ongoing)? For example, life event checklists typically sample social experiences broadly, include positive and negative items, major and minor events, and do not distinguish acute from chronic stress. In contrast, interview-based measures commonly focus on experiences that are relatively severe, include primarily negative events, and do differentiate between acute and chronic stress.<sup>3</sup> These definitional differences reflect underlying *theoretical* premises (either implicit and explicit) about the nature of life experiences assumed essen-

tial for conceptualizing stress and its consequences.<sup>4,5</sup>

Stage 2 is the operational stage. Essentially, stage 2 determines if a person's experiences fit the formal definition of life stress established at stage 1. First, does the experience meet criteria to qualify as a stressor? Second, does the experience meet criteria for particular classes of events or stress (e.g. an acute life event or a chronic stressor)? The goal of this stage is to standardize information in a consistent manner. This is a deceptively complex task. For example, a common item on self-report checklists — 'serious illness of a close family member' — may at first appear relatively straightforward. Yet 'serious' and 'close family member' are not very precise terms. How serious is 'serious', how close is 'close', and what constitutes a 'family member'? There are many instances such as these on typical life event inventories (e.g. arguments, changes at work, changes in relationships, etc.) that are not operationally explicit. Such ambiguity can result in considerable variability of interpretation across subjects.<sup>6-8</sup>

Self-report checklists attempt to constrain this variability through increasing the clarity of item wording. However, these approaches still rely on the subject's interpretation of the description on the inventory. That is, there is no information provided beyond the brief wording of the item to help ensure standardization in interpretation across individuals. In contrast, some interview-based procedures depend upon the *interviewer's or trained rater's* interpretation of the item in conjunction with the subject's report of his or her experiences to determine if the experiences qualify as a stressor. Ideally, these decisions are grounded in rules and operational criteria for translating experiences into distinctive categories of events or difficulties. However, interview-based procedures are not consistent in their approach to this issue. At one extreme, extensive rules, criteria, and case exemplars are



used to standardize raters' decisions.<sup>6,9</sup> At the other extreme, the responsibility for the interpretation is still left to the subject, who simply reports verbally to the interviewer rather than responding on the checklist.<sup>3</sup> Overall, stage 2 captures many of the pivotal points underlying the debate between self-report and interview-based methods for assessing life stress. Yet as can be seen through the above analysis, it is not necessarily an 'either/or' issue between the two methodologies. Rather, the underlying question concerns how information derived from methods that employ standardized rules and operational criteria differs from information derived from methods that do not employ such rules and criteria.

The final stage of the assessment is the quantification of the information. How is the information, as defined by stage 1 and operationalized by stage 2, combined? This quantification stage is influenced by underlying theoretical assumptions regarding the nature of stress. For example, some procedures commonly use additive measures of life events, either summing events for a total or adding events weighted by various dimensional characteristics. This procedure is based on the view that experiences sum together in producing stress-related susceptibility.<sup>10</sup> In contrast, other procedures view stress as an all-or-none issue.<sup>3,10</sup> These procedures differentiate people with and without specific forms of severe stress. Many other possibilities exist depending upon the theoretical predilection of the investigator (e.g. including only specific classes of events, such as those involving loss).

At the present stage of knowledge, it may be more productive to clarify differences in these 'inner workings' of the self-report and interview-based methods rather than to compare global features of the methods. Systematic description and classification are the heart of any science, which suggests that the basic definitional and classification procedures should be addressed first (i.e. stages 1 and 2). Once these are examined systematically, theoretical considerations can be more effectively tested (e.g. the optimal quantification of experiences — stage 3) and further refinements in theory undertaken (e.g. modifying definitions of stress and operational criteria — stages 1 and 2). The purpose of the present study is to examine: (1) how stage 1 and stage 2 information obtained with a self-report checklist differs from information based upon an interview-based procedure utilizing explicit rules and operational criteria (a 'between' method' comparison); and (2) how stage 2 information obtained

with a self-report checklist differs from information obtained by the same method as assessed six weeks later (a 'within-method' comparison).

## METHOD

### Subjects

Subjects were 92 patients presenting at Western Psychiatric Institute and Clinic at the University of Pittsburgh for treatment of a major depressive episode. All participants had a history of at least two prior major depressions, with an average of six episodes, and were participating in a treatment study of recurrent depression.<sup>11</sup> Mean age at the beginning of the protocol was 40.3 years ( $SD = 10.02$ ). Seventy-three per cent of the subjects were women, which is similar to the distribution in the general population and other clinical samples.<sup>12</sup> Ninety-six per cent of the subjects were white. Further demographic information and clinical characteristics are provided in Table 2.

Table 2 — Demographic information and clinical characteristics

Marital status	Per cent of subjects	
Married	46.7	
Divorced	25.0	
Never married	13.0	
Remarried	12.0	
Separated	3.3	
<i>Education</i>		
Less than 12 years	2.0	
High school graduate	21.0	
Some college/technical training	30.8	
College graduate	23.1	
Some graduate or professional	13.1	
<i>Employment</i>		
Full-time	49.4	
Homemaker	20.9	
Part time	14.3	
Unemployed/laid off	6.6	
Other	8.8	
<i>Clinical characteristics</i>	<i>Mean</i>	<i>SD</i>
Number of prior Episodes. (min. = 2)	6.08	6.27
HRS score at intake	22.02	4.80
BDI score at intake	28.01	7.70

### Measures

*The Psychiatric Epidemiology Research Inventory (PERI) Life Events Scale.*<sup>15</sup> The PERI is a self-report questionnaire used to assess life events. The present study used a modified 110-item version, where subjects were asked to report whether any listed event had occurred in the 12 weeks prior to the assessment date (i.e. prior to treatment entry). Subjects were also instructed to write in any events that occurred in their life that were not listed on the PERI. Each endorsed event was also rated by the patient on a seven-point scale (i.e. 1 — 'Not at all undesirable', through 7 — 'Extremely undesirable').

*Follow-up interview and ratings.* After the patient completed the PERI, he or she was then interviewed about the endorsed experiences. Specific probes were tailored to each event, and further information was sought to elaborate the possible implications of the experience for the particular individual. After the interview was completed, the information was presented by the interviewer to a panel of trained raters.

The system used for rating was predicated on the Bedford College Life Events and Difficulties Schedule (LEDS) and the extensive manuals of criteria and examples developed by Brown and Harris. (Raters were trained by SMM, who in turn received training in the LEDS procedures from one of its authors, Tirril Harris.) The LEDS system divides life experiences into events and difficulties. Events are defined as discrete stressors that last less than two weeks. The loss of a job or a fight with a spouse represent common events. Difficulties are ongoing (over a month) stressors that impact the respondent's life. For example, the event of losing a job can lead to a financial difficulty if the individual stays unemployed.

The use of specific criteria and guidelines for defining life stress follows the same logic as a structured diagnostic interview. Raters used information from the interview to decide, using explicit guidelines, whether or not the endorsed experience qualified in the system as an event or a difficulty. Specifically, the interviewer presented the information derived from the interview to the panel of raters. The panel then arrived at a consensus rating of the stressor using specific LEDS criteria to determine the type (event versus difficulty), severity, and so on.

### Procedure

The contrasting information from the initial PERI assessment and the follow-up interview assessment yielded the definitional discrepancies. There were six categories of discrepancies: (1) events that were unreported on the checklist but elicited by the interview; (2) ongoing difficulties reported as events on the checklist; (3) events reported in redundant manner on the checklist (e.g. endorsing both 'trouble at work' and 'trouble with boss' on the inventory, while the interview revealed that the subject was referring by both events to the same experience); (4) events that were mislabeled on the checklist (e.g. reporting 'quit work' when the interview revealed the subject actually was fired); (5) items reported on the checklist that did not achieve LEDS criteria for an event (e.g. a serious illness that was actually not serious by LEDS criteria); and (6) events that had no counterpart on the checklist and had to be written in.

The subjects also received a second assessment six weeks later ( $M = 41.7$  days). Life events were reported in a manner identical to the first assessment (i.e. they received the PERI and follow-up interview), except that the assessment covered the period of the prior 18 weeks. When there were inconsistencies between the two PERI reports for the 12-week period common to both assessments, the patient was directly questioned to resolve the issue. Inconsistencies included: (1) items reported at one assessment but not the other; (2) items reported at both assessments, but labeled differently (e.g. 'trouble at work' reported at the first assessment versus 'fight with a boss' reported at the second assessment); and (3) items reported as the same event at each assessment but which were actually different events (e.g. reporting argument with close friend at each assessment, but actually referring to two or more different arguments with two or more different friends).

### RESULTS

There was a significant relationship between the number of LEDS events and the number of PERI events reported by subjects ( $r = 0.41, p < 0.0001$ ). However, the number of PERI events a subject endorsed accounted for only 16 per cent of the variance in the number of LEDS events reported on the interview. This indicates a considerable discrepancy between the two measures (Fig. 1), which



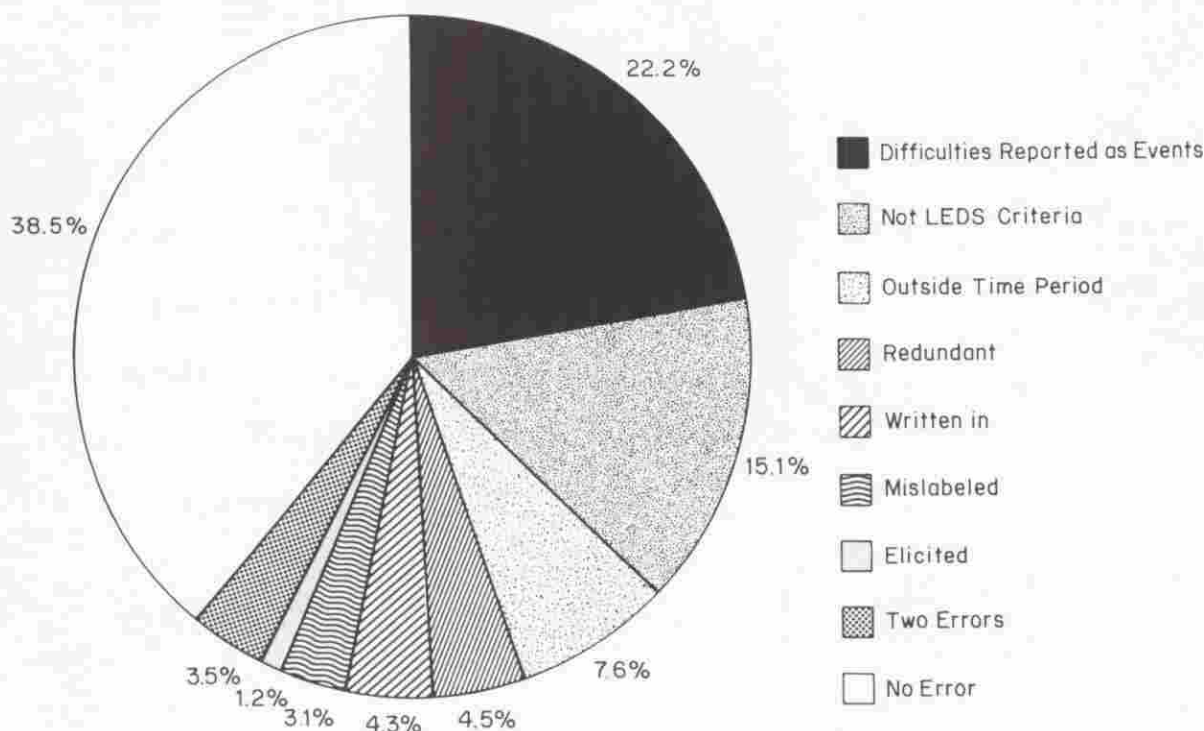


Fig. 1 — Percentage of life events that were discrepant compared with the LEDS ratings

gives added justification to detailing the sources of the discrepancy.

Sixty-two per cent of all events elicited through the checklist procedure were found to be discrepant compared to the LEDS system (314 of 510 events). There was an average of 3.6 discrepancies per subject. The most common definitional discrepancy was ongoing difficulties that were encoded by the checklist as events (36 per cent of all discrepancies). A further 25 per cent of the discrepancies were experiences that the patient reported as events that did not meet LEDS criteria. An additional 12 per cent of the discrepancies were events reported that actually were found to be outside of the designated time period. Finally, 7 per cent of the events were redundant with another event, 7 per cent were written in, 5 per cent were mislabeled, 2 per cent were unreported but elicited by interview, and 6 per cent were actually two errors (eg mislabeled and redundant with another event).

Over 60 per cent of events reported on the PERI at the first assessment were inconsistent with those events reported on the PERI at the second assessment (Fig. 2). One hundred and ninety-eight events (38.8 per cent of all events) were reported at the

first assessment but not at the second, and 52 events (10.2 per cent) were reported at the second assessment but not at the first. Six per cent were reported at both time periods, but labeled differently at each time point. Finally, 3.7 per cent were reported at both assessments but were shown on further interview to represent two or more different events. Overall, subjects had an average of 3.5 inconsistencies each.

Males differed from females on the occurrence of several types of errors and inconsistencies. Males were significantly more likely to report the same event in two different categories (Fisher's exact test;  $p < 0.001$ ). Men were also more likely to report an event at the first assessment but not the second (Fisher's exact test;  $p < 0.05$ ), use different labels for the same event at two different assessments (Fisher's exact test;  $p < 0.05$ ), or use the same label for two different events at the two assessments than females (Fisher's exact test;  $p < 0.05$ ). While women had an average of 2.8 inconsistencies and 3.2 errors per person, males had even more. Males averaged 5.3 inconsistencies each ( $t = 2.5$ ,  $p < 0.01$ ) and there was a trend for males to have more errors ( $t = 1.9$ ,  $p < 0.06$ ,  $M_s = 4.7$  vs 3.2).

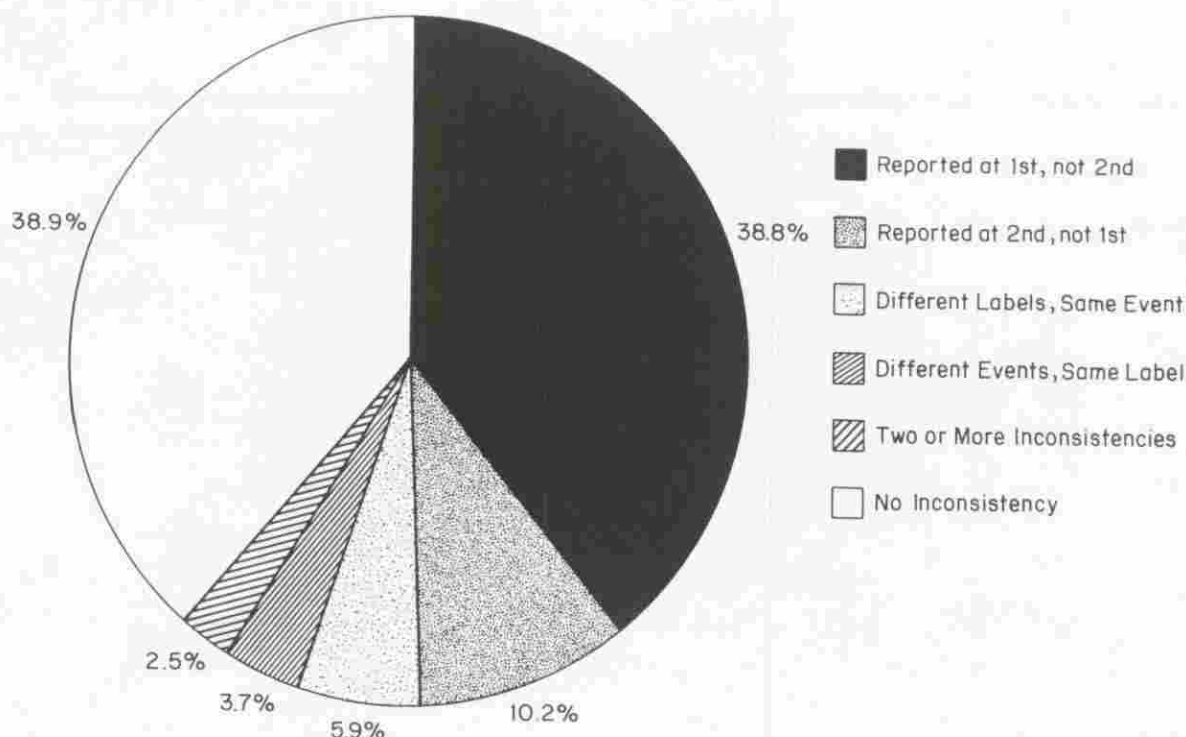


Fig. 2 — Percentage of life events that were inconsistent

### DISCUSSION

The most common approach to comparing life stress measures is to examine the correlations of the two measures. Doing so, we find some relationship between the LEDS and the PERI. However, variance accounted for was only 16 per cent, which suggests that these measures are assessing related, but by no means identical, phenomena. To understand the differences between the measures, a more in-depth examination of the differences is therefore necessary.

The translation of ongoing life experiences into categories of life events and chronic stressors represents a deceptively difficult undertaking. This procedure lies at the core of the assessment, and the sources of unreliability associated with this stage of measurement must be understood before theoretically pertinent questions can be adequately addressed. Overall, subjects reported 5.5 events on the checklist; of these, 3.6 were discrepant with the LEDS events and 3.5 were inconsistent with the follow-up checklist assessment. Other recent research has produced similar overall conclusions.<sup>7</sup> Yet to our knowledge, no studies to date have

detailed the inner workings of the two approaches and the specific aspects along which they differ (i.e. acute versus chronic, different definitional criteria, outside time period). That is, although overall outcomes have been compared, the preceding stages of definition and operationalization have not been systematically examined.

Many of the discrepancies between the two systems reflect relative disagreements rather than absolute errors. For example, the largest discrepancy was in the definition of an experience as an acute event versus a chronic difficulty (36 per cent of the events were actually chronic stressors). This issue is one that reflects back on theoretical considerations, for it is likely that acute and chronic stress have different implications for particular outcomes.<sup>8,14</sup> For example, Monroe *et al.*<sup>15</sup> found that severe events predicted time to treatment entry for depression, but severe difficulties were significant only in their interaction with severe events. Such specificity of stressor effects cannot be detected if the measurement system is not sensitive to the distinctions between the two forms of stress. The present findings suggest that perhaps over a third of



the stressors people report on self-report checklists are actually chronic in nature.

The second largest discrepancy between the self-report and interview-based methods was in the basic definition of the experience. One-quarter of the events reported on the checklist did not meet criteria for inclusion by the interview and LED-based assessment. There are two ways in which this discrepancy can be viewed. It could be simply a matter of definition. Any nomenclature for defining and classifying stress initially will be, to some degree, arbitrary; when two such systems are compared, the differences that arise also will be, to some degree, arbitrary. From this viewpoint, the investigator must decide which method is preferred for the particular purposes of the research. Prior work tends to adopt such an interpretation of these method comparisons.<sup>1,2</sup>

Another interpretation of the findings on differences in definition, though, cuts to the heart of the measurement issue. This suggests that the differences are not arbitrary, but rather that the self-report procedure yields highly inconsistent classification of experiences across individuals. Basically, current self-report procedures are vulnerable to idiosyncratic response biases.<sup>6</sup> The interview-based method, in contrast, provides a more conservative and standardized picture of the patients' stressors. Furthermore, this interpretation is bolstered by the data from this study on within-system inconsistencies for the checklist. These latter findings indicate that subjects do not demonstrate good test-retest reliability; in particular, they may use different definitions and/or criteria for their experiences at two different time points. Overall, these findings indicate considerable between- and within-subject variance in the basic definition of stress. Most importantly, these are among the first findings to specifically detail the nature of the discrepancies. Such information is essential for refining the procedures for future research.

Other comparisons between the two assessment approaches suggest more clear-cut deficiencies of the self-report method. For example, the self-report procedure is vulnerable to inaccurate dating, redundancy in reporting events, and underreporting of events. Such deficiencies represent a considerable proportion of the variance in the checklist life stress assessment. For example, totalling these types of errors indicates that over one-fourth of the events reported on the checklist for the present sample are truly inaccurate. This estimate approaches one-half of the reported events if events

that were misclassified, written in, or did not meet definitional criteria are included in what constitutes an error. These problems are underscored by the inconsistency data for the self-report checklist over time, which indicate that a number of events are quickly forgotten and go unreported (i.e. over one-third). Fewer, but still over 10 per cent, events are not reported on the first assessment but emerge at the second one. Interestingly, subjects also are inconsistent in the categories to which they assign their experiences, suggesting caution for research based on checklist methods which examine specific areas of stress in people's lives (e.g. family or work stressors).

There may be many reasons for these discrepancies and inconsistencies between the two methods. For example, our impression was that study participants wanted to be good subjects. They felt their task was to provide information, and consequently they often 'stretched' their experiences to fit the categories of the checklist. Thus, subjects would often lower the 'threshold' for events such as serious illnesses or deaths associated with close relatives or friends; upon interview, these frequently turned out to be events of distant relatives or casual friends (or even friends of relatives or friends of friends). In other words, subjects appeared to be responding in part to the perceived demand characteristics of the situation. This was particularly true, perhaps, when there were few obvious stressors to report, and they felt obligated to produce something for our efforts. While this is helpful in that it yields greater information, it again underscores the need for mechanisms to control for such gratuitous and often extraneous details.

The sex differences on reporting errors highlight other concerns. In general, males had more inconsistencies than females in their checklist responses. If this pattern holds in the general population, or in other populations in which males are more representative (such as alcohol abusers), then these groups can be expected to be even more discrepant than the sample described here. Even so, the average of 2.8 inconsistencies and 3.2 errors for women is still sufficient cause to be very concerned about using checklist measures.

A possible criticism of these findings is that depressive individuals may be more inconsistent and discrepant in their reporting than other populations, which in turn may limit the generalizability of the findings. Studies that have looked specifically at the reporting of life events have not demonstrated this to be the case. Cohen *et al.*<sup>16</sup> found

that subjects in an induced depressed state and normal subjects did not differ on the number of life events reported (although subjects in an induced elated state reported fewer negative events). Lakey and Heller<sup>17</sup> had subjects, as well as a friend of the subject, fill out a checklist regarding the subject's life events. They found that individuals with depressive symptoms were more accurate at reporting life events. However, there may still be reason to be concerned about a possible response bias. Blaney<sup>18</sup> reviewed the literature on both mood-congruent and state-dependent memory, and concluded that individuals are more likely to recall memories that are congruent with their current mood. In general, positive memories are more accessible in positive moods and negative memories are more accessible in negative moods. This could have a significant impact on the form of discrepancy that is demonstrated by an individual. Interestingly, Cohen and colleagues argued for the use of standardized investigator type measures, such as the LEDS, to control for mood-dependent response bias.

There are also advantages to using a population of depressed patients. First, a great deal of life events research is done with depressed people, and is done using life event checklists. It is important for investigators to know that they are likely to respond in an inaccurate or discrepant manner on such measures. The present study is directly generalizable to such research. Second, a main focus of this article is the delineation of the types of errors that are introduced by using a method that leaves the operationalization of stress in the hands of the respondent (i.e. the checklist approach). We have demonstrated that at least one common research population — depressed people — demonstrates questionable reliability when reporting with life events checklists. The question of whether other populations will demonstrate similar problems is an empirical one, and represents the next logical step in the methodological study of life events research. Overall, the present findings provide a strong empirical basis for seriously questioning the viability of traditional self-report checklists and a more detailed explication of why such procedures may be scientifically suspect.

Can self-report procedures be modified to correct these deficiencies? Greater clarity of wording may help with some of the problems we have detailed,<sup>7</sup> and perhaps chronic stressors can be differentiated crudely from acute stressors if respondents are provided better guidelines. Yet more importantly, it

should be emphasized that stage 2 — the translation of experiences into categories — is a subtle and complex process. This is perhaps because the subjective state of stress seems to most people, investigators included, to be so readily apparent and easily traceable to environmental origins.<sup>8</sup> Standardization of this operational procedure, though, is not easily attained. Consistent *a priori* specification of many aspects of stress is required (e.g. criteria and thresholds for inclusion and exclusion of experiences; guidelines for differentiating high-frequency events from difficulties; procedures for determining whether related events count as one or more separate events; the time frame over which continuing and evolving events are combined or distinguished) to attain coherence of the input information into categories of stress. (These concerns may be even more essential for defining chronic stressors, and for differentiating such forms of stress from the daily experiences and events that they inevitably entail.) Investigators have sufficient problems in attaining a reasonable consensus on the way such translation should proceed; leaving such complex decisions to the subject seems to be a denial of the problems and a seeking of refuge in pseudostandardization.<sup>9</sup>

The present study explicitly highlights the definitional and operational problems of the self-report checklist. But the design of the study is biased in that it does not provide a direct test of the adequacy of interview-based methods. This does not negate the serious concerns found for discrepancies, inconsistencies, and errors. It does, though, leave some broader questions of method adequacy unanswered. There are three concerns in particular. First, there is likely great variability across interview-based methods. Although some interview-based measures have an established record of psychometric adequacy (e.g. the LEDS), others may suffer from similar problems associated with the checklist procedure.<sup>1,2</sup> The core issue probably is not the method *per se* (i.e. interview or self-report), but rather the degree to which the system explicitly details how life experiences are to be systematically treated at stage 2 of the measurement process. Second, by design the information gathered at stage 1 was confined mostly to what the subject volunteered in the initial response to the checklist (although probes sometimes led to further information). Semistructured interviews, such as the LEDS, cover an extended series of questions and topics. The comprehensiveness of such approaches along with the attendant rapport suggest that more



topics previously forgotten or of great intimacy or embarrassment may be more readily uncovered.<sup>18</sup> This implies that the discrepancies reported may be an underestimate of the true differences between the two approaches. Finally, our assessments were based on depressed patients. It might be argued that these people either have more stressful lives than others or have greater difficulty reporting reliably. These points may be true, yet as we have already noted it is for these very reasons that depressed patients represent an ideal group for examining in greater detail the problems with the different measurement systems. Not only may the problems be more readily apparent, but it is also just such groups of distressed persons that are often of interest to life stress investigators.

Future research has several paths to take. In the short term, extension of this work with other populations, both clinical and non-clinical, can help more clearly delineate the generalizability of these results. For example, mood effects on life stress reporting can be examined using an approach similar to ours comparing both depressed and non-depressed individuals. Particular attention should be paid to the role of gender in detailing problems with self-report procedures. The more far-reaching follow-up to this study can take two forms. First, researchers may attempt to improve the control over definition and operationalization in checklist measures. However, as mentioned above, it is doubtful if these stages of assessment can be controlled well enough by a checklist to provide a truly reliable instrument. Yet such efforts might at least improve on current procedures, and be useful for studies in which life stress is not the primary focus or in which labor-intensive assessments are not feasible. Second, researchers may expand the use and development of investigator-based ratings. The LEDS has been developed over nearly 20 years and represents a gold standard against which other measures can be compared.<sup>8</sup> Nonetheless, such procedures are time-consuming, costly approaches to assessment that require a significant amount of training to administer. Efforts to reduce the time and costs of such procedures, without sacrificing reliability or validity, may make investigator-based procedures applicable to a broader range of research.

Overall, the present findings indicate that it is misleading at the current stage of measurement development to directly compare global scores from interview-based and checklist approaches in the assessment of life stress. Too many component

features of the approaches differ, rendering such general tests uninformative. Research on life stress needs to focus on the standardization of the operational procedures used to translate individual experiences into categories of stress (i.e. stage 2). The lack of operationalization control with checklist procedures suggests that their reliability is seriously suspect. Methodology which provides the researcher with control over the operationalization of life stress (such as the LEDS) can benefit researchers interested in assessing life stress. Only with consistency in such methods can a nomenclature of stress be developed that, with successive research, can be refined over time to capture the essential qualities of the construct and its implications for health and well-being.

## ACKNOWLEDGEMENTS

We thank Joan Anderson for ratings, data entry, data management, and data analysis; and we thank Maria Penati for assistance in the early phases of data transcription. This work was sponsored in part by National Institute of Mental Health Research Grants MH39139-6 (Scott M. Monroe) and MH29618-13 (Ellen Frank).

## REFERENCES

1. Zimmerman, M., Pfuhl, B. and Stangle, D. Life events assessment of depressed patients: A comparison of self-report and interview formats. *J. Hum. Stress* 1986; **111**: 13-19.
2. Oei, T. I. and Zwart, F. M. The assessment of life events: Self-administered questionnaire versus interview. *J. Affect. Disord.* 1986; **10**: 185-190.
3. Katschnig, H. Measuring life stress — a comparison of the checklist and the panel technique. In: *Life Events and Psychiatric Disorders: Controversial Issues*. Katschnig, H. (Ed.) Cambridge University Press, Cambridge, 1986, pp. 74-106.
4. Dohrenwend, B. P. Problems in defining and sampling the relevant population of stressful life events. In: *Stressful Life Events: Their Nature and Effects*. Dohrenwend, B. S. and Dohrenwend, B. P. (Eds) Wiley, New York, 1974, pp. 275-310.
5. Dohrenwend, B. S. and Dohrenwend, B. P. Overview and perspective for research on stressful life events. In: *Stressful Life Events: Their Nature and Effects*. Dohrenwend, B. S. and Dohrenwend, B. P. (Eds) Wiley, New York, 1974, pp. 313-331.
6. Brown, G. W. Life events, psychiatric disorder and physical illness. *J. Psychosom. Res.* 1981; **25**: 461-473.

7. Dohrenwend, B. P., Link, B. G., Kern, R., Shrout, P. E. and Markowitz, J. Measuring life events: The problem of variability within event categories. *Stress Med.* 1990; **6**: 179-188.
8. Monroe, S. M. and Roberts, J. R. Conceptualizing and measuring life stress: Problems, principles, procedures, progress. *Stress Med.* 1990; **6**: 209-216.
9. Brown, G. W. and Harris, T. O. *Social Origins of Depression*. Tavistock, London, 1978.
10. Monroe, S. M. Life events assessment: Current practices, emerging trends. *Clin. Psychol. Rev.* 1982; **2**: 435-453.
11. Frank, E. and Kupfer, D. J. Maintenance treatment of recurrent unipolar depression: Pharmacology and psychotherapy. In: *Chronic Treatments in Neuropsychiatry*. Kemali, D. and Racagni, G. (Eds) Raven Press, New York, 1985.
12. Nolen-Hoeksema, S. Sex differences on unipolar depression: Evidence and theory. *Psychol. Bull.* 1987; **101**: 259-282.
13. Dohrenwend, B. S., Krasnoff, L., Askenasy, A. R. and Dohrenwend, B. P. Exemplification of a method for scaling life events: The PERI life events scale. *J. Health Soc. Behav.* 1978; **19**: 205-229.
14. Depue, R. A. and Monroe, S. M. Conceptualization and measurement of human disorder in life stress research: The problem of chronic disturbance. *Psychol. Bull.* 1986; **99**: 36-51.
15. Monroe, S. M., Simons, A. D. and Thase, M. E. Onset of depression and time to treatment entry: Roles of life stress. *J. Consult. Clin. Psychol.* 1991; **59**: 566-573.
16. Cohen, L. H., Towbes, L. C. and Flocco, R. Effects of induced mood on self-reported life events and perceived and received social support. *J. Personal. Soc. Psychol.* 1988; **55**: 669-674.
17. Lakey, B. and Heller, K. Response biases and the relation between negative life events and psychological symptoms. *J. Personal. Soc. Psychol.* 1985; **49**: 1662-1668.
18. Blaney, P. H. Affect and memory: A review. *Psychol. Bull.* 1986; **99**: 229-246.
19. Brown, G. W. Life events and measurement. In: *Life Events and Illness*. Brown, G. W. and Harris, T. O. (Eds) Guilford, New York, 1989, pp. 3-45.



Copyright of Stress Medicine is the property of John Wiley & Sons Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.